THE UNIVERSITY *of* EDINBURGH

Extended Stochastic Dynamics: Theory, Algorithms, and Applications in Multiscale Modelling and Data Science

Xiaocheng Shang

Doctor of Philosophy
University of Edinburgh
2015

Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text.

(*Xiaocheng Shang*)

This thesis is dedicated to my parents.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, Professor Ben Leimkuhler, FRSE, for his continuous support of my PhD study and my related research: for his patience, motivation, and immense knowledge. Ben has guided me right the way through my research and during the writing of my thesis. I can not imagine having a better supervisor and mentor for my PhD study.

Besides my supervisor, I would like to thank my examiners: Ben Goddard and Tony Shardlow, for their insightful comments and encouragement, which substantially contributed to the presentation of the thesis.

My sincere thanks also goes to Assyr Abdulle, Michael Allen, Pep Español, Dmitry Fedosov, George Karniadakis, Charles Matthews, Michael Seaton, Amos Storkey, Gabriel Stoltz, Zhanxing Zhu, Konstantinos Zygalakis, and anonymous referees, all of who provided me valuable suggestions and comments for my work.

I thank my fellow students/postdocs in the group: Charles Matthews, Matthias Sachs, Janis Bajars, Gianpaolo Gobbo, and Ralf Banisch, for the stimulating discussions and fun we have had in the last three years. Also I would like to thank all the friends that I have met during my PhD, not only in Edinburgh but also in numerous schools/workshops/conferences elsewhere (I had a really great time in Zaragoza!). In particular, I would like to thank Rory Mills for proofreading parts of my thesis. I have really enjoyed playing badminton with Chris Palmer, from being easily hammered to sometimes edging a "famous win". Not to mention nice hiking trips to the Pentlands and Clatteringshaws Loch with Carmen Li. I have also really enjoyed the regular hotpot get-togethers with Yaqing Gou and Yue Hu (looking forward to the next one!).

Last but not least, I would like to thank my parents, my sister, my brother-in-law, and my partner Heng Lu, for spiritually supporting me throughout the writing of this thesis and throughout my life in general. I also gratefully acknowledge the financial support from the University of Edinburgh and China Scholarship Council.

Lay Summary

In order to gain insight into how a system behaves at the microscopic level, atomistic simulation techniques (e.g., molecular dynamics) are often adopted. In classical molecular dynamics (MD), where energy is conserved, the movement of atoms is governed by Newton's law typically associated with a semiempirical potential energy function, minus gradient of which is the conservative force. However, in real laboratories, it is not appropriate to measure the energy of the system, not only will it vary as the system size changes but also the energy is expected to have an exchange with the environment. Thus one may instead want to perturb the system by adding to its interactions an associated "thermal bath" frequently termed "thermostat"in such a way that a constant system temperature is maintained.

The goal of atomistic simulation techniques, including MD, is to reconstruct the molecular structure and chemical/physical properties of the system. In MD, this is achieved by extracting statistics from an approximation of the evolution of the system with the help of computer simulations. Due to the high frequency of bond vibrations, the timesteps accessible in numerical simulations are greatly limited to femtoseconds (a femtosecond is 10^{-15} seconds) in order to not violate the stability of the system. As a consequence, the total simulation times accessible for some real applications, even utilizing the most powerful computing resources available, are merely of milliseconds. However, in many applications including biological systems and soft matter physics (e.g., DNA, blood flow), long-time simulations are often required to better predict the macroscopic behavior of the system.

In practice, if only macroscopic properties are of interest, it may not be that necessary to take into account all the details of atoms at the microscopic level. Instead, it may be more desirable to "coarse grain" groups of atoms to form larger particles (i.e., neglecting high frequency vibrations), interacting with soft potentials, such that larger time and length scales are accessible. This thesis is concerned with one class of these coarse-graining methods, focusing on the optimal design of numerical schemes in order to achieve a substantial enhancement in numerical efficiency without losing accuracy.

Furthermore, this thesis addresses the computational benefits of novel approaches in emerging large-scale machine learning applications (i.e., "big data") that utilize a random and much smaller subset to approximate quantities of interest based on the entire dataset, thereby achieving substantial improvements in numerical efficiency. These approaches in data science can be thought of as analogues of the coarse-graining techniques in MD.

Abstract

This thesis addresses the sampling problem in a high-dimensional space, i.e., the computation of averages with respect to a defined probability density that is a function of many variables. Such sampling problems arise in many application areas, including molecular dynamics, multiscale models, and Bayesian sampling techniques used in emerging machine learning applications. Of particular interest are thermostat techniques, in the setting of a stochastic-dynamical system, that preserve the canonical Gibbs ensemble defined by an exponentiated energy function. In this thesis we explore theory, algorithms, and numerous applications in this setting.

We begin by comparing numerical methods for particle-based models. The class of methods considered includes dissipative particle dynamics (DPD) as well as a newly proposed stochastic pairwise Nosé–Hoover–Langevin (PNHL) method. Splitting methods are developed and studied in terms of their thermodynamic accuracy, two-point correlation functions, and convergence. When computational efficiency is measured by the ratio of thermodynamic accuracy to CPU time, we report significant advantages in simulation for the PNHL method compared to popular alternative schemes in the low-friction regime, without degradation of convergence rate.

We propose a pairwise adaptive Langevin (PAdL) thermostat that fully captures the dynamics of DPD and thus can be directly applied in the setting of momentum-conserving simulation. These methods are potentially valuable for nonequilibrium simulation of physical systems. We again report substantial improvements in both equilibrium and nonequilibrium simulations compared to popular schemes in the literature. We also discuss the proper treatment of the Lees–Edwards boundary conditions, an essential part of modelling shear flow.

We also study numerical methods for sampling probability measures in high dimension where the underlying model is only approximately identified with a gradient system. These methods are important in multiscale modelling and in the design of new machine learning algorithms for inference and parameterization for large datasets, challenges which are increasingly important in "big data" applications. In addition to providing a more comprehensive discussion of the foundations of these methods, we propose a new numerical method for the adaptive Langevin/stochastic gradient Nosé–Hoover thermostat that achieves a dramatic improvement in numerical efficiency over the most popular stochastic gradient methods reported in the literature. We demonstrate that the newly established method inherits a superconvergence property (fourth order convergence to the invariant measure for configurational quantities) recently demonstrated in

the setting of Langevin dynamics.

Furthermore, we propose a covariance-controlled adaptive Langevin (CCAdL) thermostat that can effectively dissipate parameter-dependent noise while maintaining a desired target distribution. The proposed method achieves a substantial speedup over popular alternative schemes for large-scale machine learning applications.

Contents

Chapter 1

Introduction

Classical molecular dynamics (MD) tracks the motion of individual atoms determined by Newton's law in the microcanonical (NVE) ensemble, where energy E (i.e., the Hamiltonian of the system), as well as the total number of particles in the system, N, and the volume of the system, V, are conserved [7, 58, 66, 105]. However, constant energy is not the appropriate setting of a real-world laboratory environment. In most cases, one wishes instead to sample the canonical (NVT) ensemble, where temperature T, as an intensive variable, is conserved, by using "thermostat" techniques [58, 79].

The idea of a thermostat is to modify dynamics so that a prescribed invariant measure is sampled. There are competing aims in this type of work. For example, one may wish to perturb the underlying Newtonian dynamics minimally, so that temporal correlations are preserved, or one may be interested in sampling rare events in a system with metastable states; thus a variety of methods have been developed [101].

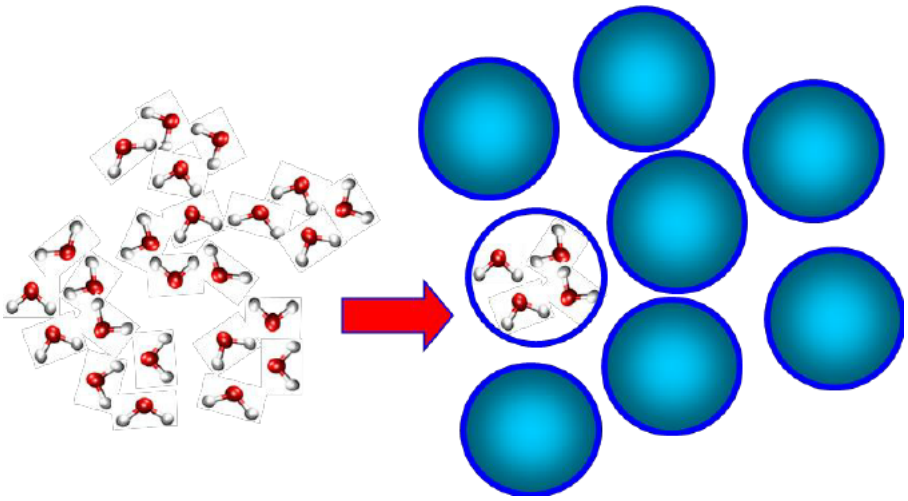


Figure 1.1: Schematic illustration of the coarse-graining procedure in DPD, where each DPD particle represents a group of four water molecules [113].

1.1 Multiscale Modelling

Stochastic momentum-conserving thermostats, which correctly capture long-ranged hydrodynamic interactions, are increasingly popular tools for simulation of simple and complex fluids [6]. The first important scheme of this type was dissipative particle dynamics (DPD), introduced by Hoogerbrugge and Koelman [74] in 1992 for simulating complex hydrodynamic behavior at a mesoscopic level that is not accessible by conventional molecular dynamics (MD) [7, 58]. In DPD, a collection of fluid molecules are grouped at the coarse-grained level and treated as a discrete particle (see Figure 1.1). These particles interact at short range in a soft potential, thereby allowing larger timesteps than would be possible in MD, while simultaneously decreasing the number of degrees of freedom required. DPD thus bridges the gap between microscale (atomistic methods, e.g., MD) and macroscale (continuum methods, e.g., Navier–Stokes) models and can be used to recover thermodynamic, dynamical, and rheological properties of complex fluids, with applications in polymer solutions [162], colloidal suspensions [132], multiphase flows [131], and biological systems [108] (see examples in Figure 1.2).

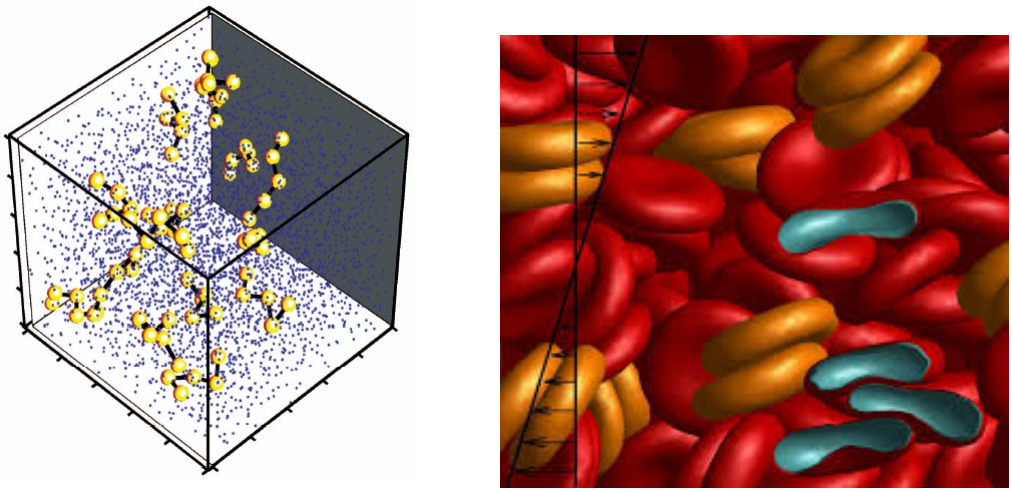


Figure 1.2: Left: Polymer chains (tethered spheres) are suspended in a solvent of DPD particles (smaller dots) [162]. Right: DPD Simulation of red blood cells under shear flow [56].

A great deal of effort has been devoted to the design of simple, efficient, and accurate numerical methods to solve the DPD system due to its promising prospects from the applications perspective. In the example of lipid bilayers, new phenomena arise as the time scale of the system that we are investigating is increased [64]. However, not all algorithms are rigorously founded and may not perform satisfactorily in large scale simulations (see discussions in [16, 31, 126, 171]). All the methods exhibit pronounced artifacts with increasing integration stepsizes due to the discretization error, typically manifest in the form of statistical bias in the calculation of thermodynamic averages. A previous study [81] suggested that, without performing serious checks for each method, the only reliable approach is to use vanishingly small stepsizes, since most of the schemes proposed are for-

mally convergent at some order of accuracy. However, we argue that using very small stepsize significantly limits the time scales accessible for DPD simulation, particularly in large scale simulations.

One of the key questions we tackle in this thesis is "What is the largest integration stepsize that can be used without damaging both static and dynamical properties?" Answering this question leads to a better understanding of the overall efficiency of each method and the validity of the schemes in the computation-intensive setting (large particle number/long-time interval). Recently, a systematic approach to thermodynamic bias in numerical computations has been used to study the accuracy and efficiency of methods for Langevin dynamics [22, 99, 100]. The approach suggested is to determine the order of accuracy of a stochastic scheme in relation to its effective invariant distribution (and, thus, with respect to steady state averages computed using numerical trajectories). This technique has led to greatly improved numerical methods for Langevin dynamics, and it is in principle applicable to momentum-conserving thermostat schemes as well. However these results are based on asymptotic expansion and thus are only relevant in the limit of small stepsize (whereas we are interested in the large stepsize threshold). Moreover we find that the analytical computations necessary to perform expansions in the DPD and pairwise Nosé–Hoover–Langevin (PNHL) [106] cases are in typical cases highly complex; we therefore restrict ourselves in this thesis to outlining some fundamental and illustrative applications of the theory. In particular for certain symmetric methods, we can demonstrate the even order approximation of long-time averages. The same conclusion may be reached for additional schemes of a specific structure (related to the geometric Langevin algorithms of [22]).

A new stochastic momentum-conserving thermostat is introduced that can be used in place of DPD in the low-friction regime or in nonequilibrium molecular dynamics (NEMD) based on stochastic extension of a scheme in a recent paper by Allen and Schmid [6]. This method is particularly inexpensive to implement and is found to have a very high stepsize stability threshold compared to alternatives.

In typical cases, for understanding the stepsize stability threshold and the performance of different schemes, we are forced to rely on numerical experiments. An excellent survey of the performance of a number of methods was undertaken by Nikunen et al. [126] in 2003. Since then, and despite many additions to the arsenal of methods, such a comparison has been lacking. To this end, we test a number of popular methods [16, 37, 155] from the DPD literature together with additional methods [114, 136, 160] that are used in popular software packages. For each method, we examine calculations such as kinetic and configurational temperatures, average potential energy, radial distribution function, velocity autocorrelation function, and transverse momentum autocorrelation function, which gives information on the rotational relaxation process.

1.2 Nonequilibrium Modelling

In equilibrium methods, transport coefficients (e.g., diffusion coefficient and shear viscosity [14]) are measured based on time integrals of correlation functions (i.e., the Green–Kubo formulas [63, 92]). The simulated system is first equilibrated and subsequently remains in equilibrium, thus simple periodic boundary conditions are adequate. For instance, the shear viscosity can be calculated by integrating the stress-stress autocorrelation function (see examples in the DPD literature [12, 31, 87]). However, time correlation functions represent the average response to the naturally occurring (hence very small) fluctuations in the system properties. Due to significant statistical error, the signal-to-noise ratio is particularly unfavorable at long times (e.g., the equilibrium stress is subject to large fluctuations in measuring the shear viscosity [12]), where there may be a significant contribution to the integral defining a transport coefficient [7, 130]. Therefore, in order to get highly converged values for the Green–Kubo integrand over the range of times which contribute significantly to the integral, the simulation must be run for long times [130]. Furthermore, the maximum time for which reliable correlations can be calculated may be limited due to the finite system size [7, 130].

For the calculation of the shear viscosity, an alternative method, which is based on the transverse-current autocorrelation function, was proposed in [130]. The method relies on an assumption of the hydrodynamic prediction for its functional form in order to extract the viscosity from the decay of this function. However, the additional assumption means that it is not obviously preferable to the Green–Kubo approach, although it is efficient for purely dissipative systems [70]. These two equilibrium methods mentioned have been compared to nonequilibrium methods in molecular dynamics simulations, showing that, despite their undoubted advantages, both equilibrium methods suffer from worse statistics than nonequilibrium methods [71].

In nonequilibrium methods, the steady state of the system is subject to external perturbations (either stationary fluxes or spatial gradients of some quantities), in order to artificially induce larger fluctuations and thus dramatically improve the signal-to-noise ratio of the measured response [51]. Therefore nonequilibrium methods typically converge more quickly than equilibrium methods. By measuring the steady state response to such a perturbation, the decay to the equilibrated state is then related to the corresponding transport coefficient, avoiding problems with long-time behavior of correlation functions [7]. For example, a periodic perturbation was employed to generate an oscillatory velocity profile that enables the measurement of transport properties of liquids by molecular dynamics simulations [33]. A more recent approach imposed a pulsed Gaussian velocity profile on the system, and an estimate of the shear viscosity was then obtained by measuring the decay of the Gaussian peak [9]. A linear velocity profile is also obtainable by cleverly swapping impulses of spatially remote particles [123]. In this case, the shear viscosity is provided with the interchanged momentum and the measured

velocities.

A different approach is often adopted in real experiments to measure the shear viscosity: a linear profile is imposed at a fixed shear rate and then the resultant shear stress can be measured. However, in computer simulations, simple periodic boundary conditions are unable to maintain a steady linear velocity profile, resulting in problems at the boundaries of the simulation domain. There have been early attempts to generate momentum or energy flows in molecular dynamics simulations where particles are made to interact with momentum or energy reservoirs (e.g., a velocity profile can be obtained by fixing the average velocity in the extremal slabs of a fluid) [11, 34, 85, 169]. However, these methods are not compatible with periodic boundary conditions, and thus lead to surface effects. Alternatively and more appealingly, one can apply the Lees–Edwards boundary conditions [94] to retain periodicity but alter the position and velocity of the periodic images. In this case, a simple shear flow (with a constant shear rate) is generated, which allows the investigation of the dependence of the viscosity on the shear rate [57] (see Figure 1.3). Lees–Edwards boundary conditions and their modifications (see discussions in [21]) have been extensively studied in DPD and related systems to study rheological behavior in colloidal suspensions [17, 18], polymeric systems [55, 133, 158], as well as multiphase systems [131] (see more discussions on boundary conditions in DPD in [138, 139, 142]).

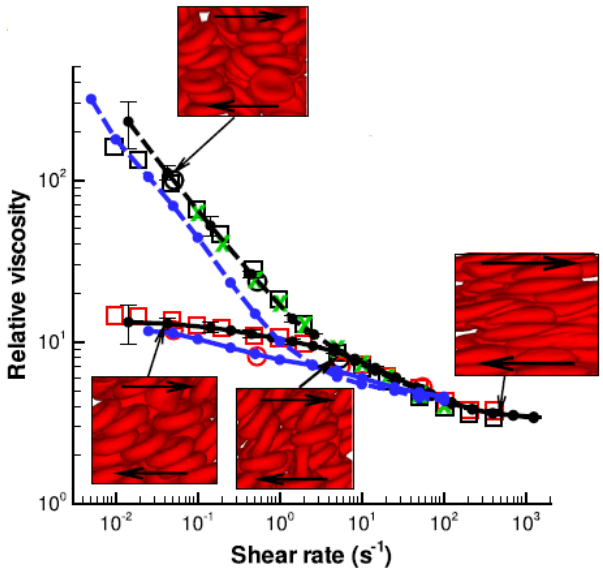


Figure 1.3: Predicting human blood viscosity in silico by applying the Lees–Edwards boundary conditions in DPD [57].

In this thesis, we focus on the practical implementation of the Lees–Edwards boundary conditions in DPD and related momentum-conserving systems. Furthermore, it should be noted that extra care is needed for any nonequilibrium models where the use of external fields would cause viscous heating (i.e., the energy pumped into the system would cause the temperature to rise monotonically for a steady perturbation) [12]. In those cases, thermostats are required to drain

the excess energy in order to maintain the correct system temperature. Thus, we also investigate the temperature control in DPD and related systems under Lees–Edwards boundary conditions, which is a nontrivial task and has been lacking in the literature [82, 133, 158].

1.3 Data Science

Monte Carlo sampling for Bayesian posterior inference is a common approach used in machine learning. For instance, hybrid Monte Carlo (HMC) methods [24, 45, 77], that have become increasingly popular recently, define a Hamiltonian function in terms of the target distribution and a "kinetic energy" based on a set of auxiliary variables termed "momenta". One then can draw states from a Hamiltonian system, utilizing its attractive properties of rapid exploration of the phase space [24, 73, 174].

However, in emerging machine learning applications, direct sampling with the entire large-scale dataset is computationally infeasible. For instance, standard Markov chain Monte Carlo (MCMC) methods [118], as well as typical HMC methods, require the calculation of the acceptance probability and the creation of informed proposals based on the whole dataset.

In order to improve the computational efficiency, a number of stochastic gradient methods [32, 44, 173, 175] have been proposed in the setting of Bayesian sampling based on random (and much smaller) subsets to approximate the likelihood of the whole dataset, thus substantially reducing the computational cost in practice. Welling and Teh proposed the so-called stochastic gradient Langevin dynamics (SGLD) [175], combining the ideas of stochastic optimization [144] and traditional Brownian dynamics, with a sequence of stepsizes decreasing to zero. A fixed stepsize is often adopted in practice which is the choice in this thesis as in Vollmer et al. [173], where a modified SGLD (mSGLD) was also introduced that was designed to reduce the sampling bias.

SGLD generates samples from first order Brownian dynamics, and thus, with a fixed timestep, one can show that it is unable to dissipate excess noise in gradient approximations while maintaining the desired invariant distribution [32]. A stochastic gradient Hamiltonian Monte Carlo (SGHMC) method was proposed by Chen et al. [32], which relies on second order Langevin dynamics and incorporates a parameter-dependent diffusion matrix that is intended to effectively offset the stochastic perturbation of the gradient. However, it is difficult to accommodate the additional diffusion term in practice. Moreover, as pointed out in [44], poor estimation of it may have a significant adverse influence on the sampling of the target distribution; for example, the effective system temperature may be altered.

The "thermostat" idea, which is widely used in molecular dynamics [58, 101], was recently adopted in the stochastic gradient Nosé–Hoover thermostat

(SGNHT) by Ding et al. [44] in order to adjust the kinetic energy during simulation in such a way that the canonical ensemble is preserved (i.e., so that a prescribed constant temperature distribution is maintained). In fact, the SGNHT method is essentially equivalent to the adaptive Langevin (Ad-Langevin) thermostat proposed earlier by Jones and Leimkuhler [84] in the molecular dynamics setting (see [107] for discussions).

Despite the substantial interest generated by these methods, the mathematical foundation for stochastic gradient methods has been incomplete. The underlying dynamics of the SGNHT method [44] was taken up in a recent work with B. Leimkuhler [107], together with the design of discretization schemes with high effective order of accuracy. SGNHT methods are designed based on the assumption of constant noise variance. We propose a covariance-controlled adaptive Langevin (CCAdL) thermostat, that can handle parameter-dependent noise, improving both robustness and reliability in practice, and which can effectively speed up the convergence to the desired invariant distribution in large-scale machine learning applications.

1.4 Outline of the Thesis

The rest of the thesis is organized as follows.

In Chapter 2, we review the formulations of a wide range of particle-based systems, ranging from deterministic Hamiltonian dynamics to various stochastic thermostats. The properties of equations of motion of those systems are discussed, followed by an introduction of two widely used boundary conditions in equilibrium and nonequilibrium molecular dynamics (i.e., the periodic and Lees–Edwards boundary conditions, respectively).

Chapter 3 concerns the numerical treatment of various methods introduced in the previous chapter, with a focus on DPD and related systems. We investigate the order of convergence to the invariant measure for those momentum-conserving thermostats. Extensive numerical experiments are then carried out to systematically study not only the order of convergence but also the performance (such as accuracy, robustness, and numerical efficiency) of those methods in practice. We also demonstrate that a nonsymmetric splitting method of PNHL has second order convergence to the invariant measure for certain observables.

A new momentum-conserving thermostat, the pairwise adaptive Langevin (PAdL) thermostat, is proposed in Chapter 4 that fully captures the dynamics (i.e., the correct decay of the velocity autocorrelation function, which is related to the translational diffusion coefficient) of the DPD system. We compare PAdL with alternative systems in both equilibrium and nonequilibrium simulations. In particular, the Lees–Edwards boundary conditions are applied to generate a simple shear flow in the nonequilibrium case. We discuss the practical imple-

mentation of the Lees–Edwards boundary conditions and then investigate the temperature control of various methods under shear flow.

In Chapter 5, we study numerical methods for noisy gradient systems, where additional noise is present that inhibits an efficient sampling of the invariant measure. Assuming a constant covariance of the noise, we provide a comprehensive discussion of the adaptive Langevin/stochastic gradient Nosé–Hoover thermostat, that can effectively dissipate excess noise while maintaining a desired invariant distribution. We also propose a covariance-controlled adaptive Langevin (CCAdL) thermostat that can deal with parameter-dependent noise.

Our findings are summarized in Chapter 6, with key contributions of this thesis highlighted.

Chapter 2

Formulations and Properties of Equations of Motion

In this chapter, we review the formulations and properties of equations of motion of the Hamiltonian dynamics and a variety of isothermal (i.e., temperature-conserving) systems, followed by an introduction of the periodic and Lees–Edwards boundary conditions that have been widely used in equilibrium and nonequilibrium molecular dynamics, respectively. In particular, the pairwise Nosé–Hoover–Langevin (PNHL) thermostat was proposed in a recent publication [106] with B. Leimkuhler.

2.1 Hamiltonian Dynamics

Consider an N-particle system evolving in dimension d with positions $\mathbf{q}_i \in \mathbb{R}^d$, momenta $\mathbf{p}_i \in \mathbb{R}^d$, and masses $m_i \in \mathbb{R}$ for $i = 1, \ldots, N$. The time evolution of the Hamiltonian system is governed by Newton's equations

$$\frac{\mathrm{d}\mathbf{q}}{\mathrm{d}t} = \mathbf{M}^{-1}\mathbf{p}\,, \tag{2.1}$$

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = -\nabla U(\mathbf{q})\,, \tag{2.2}$$

where $\mathbf{q} = \left[\mathbf{q}_1^T, \ldots, \mathbf{q}_N^T\right]^T$, $\mathbf{p} = \left[\mathbf{p}_1^T, \ldots, \mathbf{p}_N^T\right]^T \in \mathbb{R}^{dN}$, $\mathbf{M} = \mathrm{diag}(m_1\mathbf{I}_d, \ldots, m_N\mathbf{I}_d)$ is the diagonal mass matrix, and $-\nabla U(\mathbf{q}) \equiv \mathbf{F}(\mathbf{q})$ represents the conservative force. Overall, the total energy (Hamiltonian) of the system is conserved, i.e.,

$$H(\mathbf{q}, \mathbf{p}) = \frac{\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p}}{2} + U(\mathbf{q}) = \mathrm{const}\,, \tag{2.3}$$

where $U(\mathbf{q}) : \mathbb{R}^{dN} \to \mathbb{R}$ is the potential energy function. This property can be easily demonstrated by differentiating the Hamiltonian with respect to time and

using (2.1)–(2.2)

$$\frac{\mathrm{d}H(\mathbf{q},\mathbf{p})}{\mathrm{d}t} = \nabla U(\mathbf{q}) \cdot \frac{\mathrm{d}\mathbf{q}}{\mathrm{d}t} + \mathbf{M}^{-1}\mathbf{p} \cdot \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = 0\,. \tag{2.4}$$

Therefore conservation of the total energy is a common choice of testing the correctness of computer simulations of Hamiltonian dynamics [66, 105].

In a simple example of a one-dimensional harmonic oscillator with a single particle (unit mass), the potential energy function is given by

$$U(q) = \frac{1}{2}\omega^2 q^2\,, \tag{2.5}$$

where ω is a positive parameter. It can be physically interpreted as a particle connected to the origin by a linear spring with spring constant ω^2. Thus the equations of motion of the particle can be written as

$$\frac{\mathrm{d}q}{\mathrm{d}t} = p\,, \tag{2.6}$$

$$\frac{\mathrm{d}p}{\mathrm{d}t} = -\omega^2 q\,. \tag{2.7}$$

Given initial values of $q(0) = q_0$ and $p(0) = p_0$, we are in fact able to solve the system "exactly" to obtain the oscillatory solution with frequency ω:

$$\begin{bmatrix} q(t) \\ p(t) \end{bmatrix} = \begin{bmatrix} \cos(\omega t) & \omega^{-1}\sin(\omega t) \\ -\omega \sin(\omega t) & \cos(\omega t) \end{bmatrix} \begin{bmatrix} q_0 \\ p_0 \end{bmatrix}. \tag{2.8}$$

However, it should be noted that in most of the cases in molecular dynamics and other applications, the potential energy function could be much more complicated, resulting in a nonlinear system that we are not able to solve exactly. For instance, the potential energy function in molecular dynamics is usually assumed to be a semiempirical formula constructed from primitive functions via a prior parameter fitting procedure. Alternatively, one may assume that it is the probability distribution ($\rho > 0$) that is specified and that the potential energy is constructed from it via

$$U = -\beta^{-1} \ln \rho\,, \tag{2.9}$$

where β is a positive parameter and can be interpreted as being proportional to the reciprocal temperature in an associated physical system, i.e.,

$$\beta^{-1} = k_{\mathrm{B}} T\,, \tag{2.10}$$

where k_{B} is the Boltzmann constant and T is the temperature. In those cases, we are forced to use "discretization methods", which we will address in later Chapters, to approximate the propagation of the true solution.

2.2 Brownian and Langevin Dynamics

Hamiltonian dynamics has been widely used in many applications, such as classical molecular dynamics (MD) [7, 58] where energy is conserved. However, as already mentioned, constant energy is not the appropriate setting of a real-world laboratory environment. Thus, in most cases, one wishes instead to sample the canonical ensemble, where temperature is conserved, which leads to the development of various thermostat techniques [58, 79].

The most obvious proposals, and also the oldest, are Brownian and Langevin dynamics. In Brownian (sometimes called "overdamped Langevin") dynamics, the system is

$$\mathrm{d}\mathbf{q} = -\lambda \nabla U(\mathbf{q})\mathrm{d}t + \sqrt{2\beta^{-1}\lambda}\mathrm{d}\mathbf{W}\,, \tag{2.11}$$

where $\mathrm{d}\mathbf{W}$ represents a vector of infinitesimal Wiener increments, and λ is a free parameter which represents a time-rescaling.

One can write down the corresponding Fokker–Planck operator of the Brownian dynamics (BD) (2.11) as

$$\mathcal{L}^{\dagger}_{\mathrm{BD}} = \lambda \nabla_{\mathbf{q}} \cdot (\nabla U(\mathbf{q}) \cdot) + \beta^{-1}\lambda \Delta_{\mathbf{q}}\,. \tag{2.12}$$

It can be shown [28] that this system (2.11) ergodically samples the Gibbs–Boltzmann probability distribution

$$\bar{\rho}_{\beta} = Z^{-1} \exp\left(-\beta U(\mathbf{q})\right)\,, \tag{2.13}$$

where Z is a suitable normalizing constant often interpreted as the partition function. In other words, the Gibbs–Boltzmann distribution (2.13) is the invariant (stationary) distribution of the system, satisfying the following relation:

$$\mathcal{L}^{\dagger}_{\mathrm{BD}} \bar{\rho}_{\beta} = 0\,. \tag{2.14}$$

For simplicity, we assume that the configurations $\mathbf{q}$ are restricted to a compact and simply connected domain $\Omega_{\mathbf{q}}$. In many applications it is found that the use of a first order dynamics such as (2.11) is inefficient or introduces unphysical dynamical properties, and one employs, instead, the Langevin dynamics (LD) method:

$$\mathrm{d}\mathbf{q} = \mathbf{M}^{-1}\mathbf{p}\mathrm{d}t\,, \tag{2.15}$$

$$\mathrm{d}\mathbf{p} = -\nabla U(\mathbf{q})\mathrm{d}t - \gamma \mathbf{p}\mathrm{d}t + \sqrt{2\beta^{-1}\gamma}\mathbf{M}^{1/2}\mathrm{d}\mathbf{W}\,, \tag{2.16}$$

with a modified Fokker–Planck operator

$$\mathcal{L}^{\dagger}_{\mathrm{LD}} = -\mathbf{M}^{-1}\mathbf{p} \cdot \nabla_{\mathbf{q}} + \nabla U(\mathbf{q}) \cdot \nabla_{\mathbf{p}} + \gamma \nabla_{\mathbf{p}} \cdot (\mathbf{p}\cdot) + \beta^{-1}\gamma \nabla_{\mathbf{p}} \cdot (\mathbf{M}\nabla_{\mathbf{p}}\cdot)\,. \tag{2.17}$$

Again, γ in these equations is a free parameter, termed the "friction constant".

It is related to the timescale on which the variables of the system interact with particles of a fictitious extended "bath", but it cannot be associated with a simple time-rescaling of the equations of motion and is thus different from λ in (2.11). It is a little more involved to show that (2.15)–(2.16) ergodically [116] samples the distribution with density

$$\rho_\beta = Z^{-1}\exp(-\beta H(\mathbf{q},\mathbf{p}))\,, \tag{2.18}$$

i.e.,

$$\mathcal{L}_{\mathrm{LD}}^\dagger \rho_\beta = 0\,. \tag{2.19}$$

In molecular dynamics, the matrix $\mathbf{M}$ is typically diagonal and contains the masses of atoms. In more general settings, the masses and friction coefficient may be treated as free parameters, and by computing long trajectories of (2.15)–(2.16), one may obtain averages with respect to $\bar{\rho}_\beta(\mathbf{q})$; i.e., if $\{(\mathbf{q}(\tau),\mathbf{p}(\tau)) : \tau \geq 0\}$ is a path generated by solving the SDE system (2.15)–(2.16), one has, for suitable test functions $\phi(\mathbf{q})$, and under certain conditions on the potential energy function U [116],

$$\lim_{\tau\to\infty} \tau^{-1} \int_0^\tau \phi(\mathbf{q}(\tau))\,\mathrm{d}\tau = \int_{\Omega_\mathbf{q}} \phi(\mathbf{q})\bar{\rho}_\beta(\mathbf{q})\,\mathrm{d}\omega_\mathbf{q}\,,$$

where $\mathrm{d}\omega_\mathbf{q} = \mathrm{d}\mathbf{q}_1\mathrm{d}\mathbf{q}_2\ldots\mathrm{d}\mathbf{q}_N$. In other words, the projected path defines a sampler for the density $\bar{\rho}_\beta$. Langevin dynamics can thus be seen as an extended system which allows sampling to be performed in a reduced cross section of phase space by marginalization over long trajectories; this is the essential property of a thermostat.

2.3 Nosé–Hoover Thermostat

A number of thermostats have been developed for molecular dynamics simulations [79], among which one of the most popular ones is the so-called Nosé–Hoover thermostat [75, 76, 127, 128]. The equations of motion of the Nosé–Hoover thermostat can be written as

$$\begin{aligned}
\mathrm{d}\mathbf{q} &= \mathbf{M}^{-1}\mathbf{p}\mathrm{d}t\,,\\
\mathrm{d}\mathbf{p} &= -\nabla U(\mathbf{q})\mathrm{d}t - \xi\mathbf{p}\mathrm{d}t\,,\\
\mathrm{d}\xi &= \mu^{-1}\left[\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p} - N_\mathrm{d}k_\mathrm{B}T\right]\mathrm{d}t\,,
\end{aligned} \tag{2.20}$$

where μ is a coupling parameter which is referred to as the "thermal mass" in the molecular dynamics setting. The auxiliary variable $\xi \in \mathbb{R}$ is governed by a Nosé–Hoover device via a negative feedback mechanism, i.e., when the instantaneous temperature (average kinetic energy per degree of freedom) calculated as

$$k_\mathrm{B}T = \frac{\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p}}{N_\mathrm{d}} \tag{2.21}$$

is below the target temperature, the "dynamical friction" ξ would decrease allowing an increase of temperature, while ξ would increase when the temperature is above the target. It should be noted that the number of degrees of freedom of the system is typically the product of the dimension and the number of particles, i.e., $N_{\mathrm{d}} = dN$. However, when subject to constrains, the number of constrains should be subtracted from the number of degrees of freedom accordingly (see more discussions in [79]).

It has been demonstrated that the use of purely deterministic approaches, such as Nosé–Hoover, results in ergodicity issues [95, 96]. Assuming ergodicity, one can easily verify that the Nosé–Hoover system (2.20) preserves the canonical ensemble with a modified invariant distribution

$$\hat{\rho}_\beta(\mathbf{q}, \mathbf{p}, \xi) = \frac{1}{Z} \exp\left(-\beta H(\mathbf{q}, \mathbf{p})\right) \exp\left(-\frac{\beta\mu\xi^2}{2}\right). \tag{2.22}$$

Due to the fact that the distribution of the auxiliary variable ξ is separable from those of $\mathbf{q}$ and $\mathbf{p}$ (i.e., the overall density is in product form), and functions of ξ is easily normalizable, one can always average out over the auxiliary variable to compute the averages of functions of $\mathbf{q}$ and $\mathbf{p}$ as in the Langevin dynamics.

2.4 Nosé–Hoover–Langevin (NHL) Thermostat

In order to improve the ergodicity, the auxiliary variable in the Nosé–Hoover thermostat is further coupled to stochastic processes of Ornstein–Uhlenbeck type. This is known as the Nosé–Hoover–Langevin (NHL) thermostat [104, 150], which reads

$$\begin{aligned} \mathrm{d}\mathbf{q} &= \mathbf{M}^{-1}\mathbf{p}\mathrm{d}t\,, \\ \mathrm{d}\mathbf{p} &= -\nabla U(\mathbf{q})\mathrm{d}t - \xi\mathbf{p}\mathrm{d}t\,, \\ \mathrm{d}\xi &= \mu^{-1}\left[\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p} - N_{\mathrm{d}}k_{\mathrm{B}}T\right]\mathrm{d}t - \tilde{\gamma}\xi\mathrm{d}t + \tilde{\sigma}\mathrm{d}\mathrm{W}\,, \end{aligned} \tag{2.23}$$

where coefficient constants $\tilde{\gamma}$ and $\tilde{\sigma}$ satisfy the fluctuation-dissipation theorem in standard Langevin dynamics

$$\tilde{\sigma}^2 = \frac{2\tilde{\gamma}k_{\mathrm{B}}T}{\mu}\,, \tag{2.24}$$

and $\mathrm{W} = \mathrm{W}(t)$ is a standard Wiener process.

The ergodicity of the Nosé–Hoover–Langevin thermostat has been analytically proved by Leimkuhler et al. [104] in the case of a harmonic system. One can show that the Nosé–Hoover–Langevin system (2.23) also preserves the canonical ensemble with the same invariant distribution (2.22) as in Nosé–Hoover. Other types of thermostats that sample the canonical ensemble have also been proposed, e.g., the generalized Bulgac–Kusnezov methods [97]. The use of auxiliary variables can provide a degree of flexibility in the design of the thermostat, for example, allowing the treatment of systems arising in fluid dynamics [46] or imposing an

isokinetic constraint [98].

2.5 Dissipative Particle Dynamics (DPD)

The dissipative particle dynamics (DPD) method was first proposed by Hoogerbrugge and Koelman [74] in order to recover the properties of isotropy and Galilean invariance that were both broken in the so-called lattice-gas automata (LGA) method [59]. The original system itself was updated in discrete time steps and was later reformulated by Español and Warren [50] as a system of stochastic differential equations (SDEs).

The time evolution of each DPD particle $i = 1, \ldots, N$ (in dimension d), representing a cluster of molecules, is governed by Newton's equations

$$\frac{\mathrm{d}\mathbf{q}_i}{\mathrm{d}t} = \frac{\mathbf{p}_i}{m_i}\,, \quad \frac{\mathrm{d}\mathbf{p}_i}{\mathrm{d}t} = \mathbf{F}_i\,, \tag{2.25}$$

where $\mathbf{q}_i$, $\mathbf{p}_i$, and m_i denote positions, momenta, and masses, respectively. $\mathbf{F}_i$ represents the total interparticle force acting on particle i due to the presence of the other particles, which is a summation of three pairwise contributions

$$\mathbf{F}_i = \sum_{j \neq i} (\mathbf{F}_{ij}^{\mathrm{C}} + \mathbf{F}_{ij}^{\mathrm{D}} + \mathbf{F}_{ij}^{\mathrm{R}})\,, \tag{2.26}$$

where $\mathbf{F}_{ij}^{\mathrm{C}}$, $\mathbf{F}_{ij}^{\mathrm{D}}$, and $\mathbf{F}_{ij}^{\mathrm{R}}$ represent conservative, dissipative, and random forces, acting on particle i due to the j-th particle, respectively.

The conservative force that controls the thermodynamics of the DPD system is normally [65] chosen as

$$\mathbf{F}_{ij}^{\mathrm{C}} = \begin{cases} a_{ij}\left(1 - \dfrac{q_{ij}}{r_{\mathrm{c}}}\right)\hat{\mathbf{q}}_{ij}\,, & q_{ij} < r_{\mathrm{c}}\,; \\ 0\,, & q_{ij} \geq r_{\mathrm{c}}\,, \end{cases} \tag{2.27}$$

where parameter a_{ij} ($a_{ij} = a_{ji} \geq 0$) represents the maximum repulsion strength between each interacting pairs, and r_{c} denotes a cutoff radius that is often used in order to reduce the computational cost. The relative positions between interacting pairs are denoted by $\mathbf{q}_{ij} = \mathbf{q}_i - \mathbf{q}_j$ with length $q_{ij} = \|\mathbf{q}_{ij}\|$ and the unit direction from $\mathbf{q}_j$ to $\mathbf{q}_i$ by $\hat{\mathbf{q}}_{ij} = \mathbf{q}_{ij}/q_{ij}$.

The dissipative and random forces, forming a thermostat to preserve the system temperature, are given by

$$\mathbf{F}_{ij}^{\mathrm{D}} = -\gamma\omega^{\mathrm{D}}(q_{ij})(\hat{\mathbf{q}}_{ij} \cdot \mathbf{v}_{ij})\hat{\mathbf{q}}_{ij}\,, \tag{2.28}$$

$$\mathbf{F}_{ij}^{\mathrm{R}} = \sigma\omega^{\mathrm{R}}(q_{ij})\theta_{ij}\hat{\mathbf{q}}_{ij}\,, \tag{2.29}$$

where γ and σ are constant parameters representing the dissipative and random strengths respectively, ω^{D} and ω^{R} are position-dependent weight functions, and $\mathbf{v}_{ij} = \mathbf{p}_i/m_i - \mathbf{p}_j/m_j$ denote the relative velocities. θ_{ij} is a symmetric ($\theta_{ij} = \theta_{ji}$) Gaussian white-noise term with the following stochastic property

$$\langle \theta_{ij}(t) \rangle = 0\,, \qquad \langle \theta_{ij}(t)\theta_{kl}(t') \rangle = (\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})\delta(t-t')\,, \tag{2.30}$$

where the angle brackets denote the averages, δ_{ik}, δ_{jl}, δ_{il}, and δ_{jk} are the Kronecker delta functions, and $\delta(t-t')$ is the Dirac delta function, and is chosen independently for each interacting pair at each time step.

The canonical ensemble was not preserved in the original formulation of DPD [74]. This has been corrected by Español and Warren [50], which leads to the fluctuation-dissipation theorem for the DPD system, namely:

$$\omega^{\mathrm{D}}(q_{ij}) = \left[\omega^{\mathrm{R}}(q_{ij})\right]^2\,, \quad \sigma^2 = 2\gamma k_{\mathrm{B}} T\,. \tag{2.31}$$

Then it can be easily demonstrated that the canonical ensemble is preserved with an invariant distribution defined by the density ρ_β (2.18) as in Langevin dynamics.

The conservative potential energy is defined as

$$U(\mathbf{q}) = \sum_i \sum_{j>i} U(q_{ij})\,, \tag{2.32}$$

where $U(q_{ij})$ is the "soft" pair potential associated with the conservative force (2.27):

$$U(q_{ij}) = \begin{cases} \dfrac{a_{ij} r_{\mathrm{c}}}{2}\left(1 - \dfrac{q_{ij}}{r_{\mathrm{c}}}\right)^2\,, & q_{ij} < r_{\mathrm{c}}\,; \\ 0\,, & q_{ij} \geq r_{\mathrm{c}}\,. \end{cases} \tag{2.33}$$

Although we write the density (2.18) as an exponential, we note that if the total momentum is conserved, the density should be replaced by

$$\begin{aligned} \rho_\beta(\mathbf{q},\mathbf{p}) = {} & Z^{-1} \exp(-\beta H(\mathbf{q},\mathbf{p})) \\ & \times \delta\left[\sum_i p_{i,x} - \pi_x\right] \delta\left[\sum_i p_{i,y} - \pi_y\right] \delta\left[\sum_i p_{i,z} - \pi_z\right]\,, \end{aligned} \tag{2.34}$$

where $\boldsymbol{\pi} = (\pi_x, \pi_y, \pi_z)$ is the total momentum vector. A similar modification would be needed if the angular momentum were also conserved. It is worthy of mention that the ergodicity of the DPD system has only been demonstrated in the case of high particle density in one dimension by Shardlow and Yan [156].

Due to the fact that the DPD system depends on relative velocities and the interactions between particles are pairwise (symmetric), both total and angular momenta are conserved, thus DPD is an isotropic Galilean-invariant thermostat which also preserves hydrodynamics [6, 120]. However, if periodic boundary con-

ditions are used, the conservation of the angular momentum will be destroyed.

One of the two weight functions can be chosen arbitrarily without violating thermal equilibrium. A simple choice has been widely used, which reads

$$\omega^{\mathrm{R}}(q_{ij}) = \begin{cases} 1 - \dfrac{q_{ij}}{r_{\mathrm{c}}}\,, & q_{ij} < r_{\mathrm{c}}\,; \\ 0\,, & q_{ij} \geq r_{\mathrm{c}}\,. \end{cases} \tag{2.35}$$

In this case the conservative force (2.27) can be written in a compact way

$$\mathbf{F}_{ij}^{\mathrm{C}} = a_{ij}\omega^{\mathrm{R}}(q_{ij})\hat{\mathbf{q}}_{ij}\,. \tag{2.36}$$

To make the presentation simpler, a compact form the DPD system (2.25) (for particle i) may be used [6]

$$\begin{aligned} \mathrm{d}\mathbf{q}_i &= m_i^{-1}\mathbf{p}_i \mathrm{d}t\,, \\ \mathrm{d}\mathbf{p}_i &= \mathbf{F}_i^{\mathrm{C}}(\mathbf{q})\mathrm{d}t - \gamma\mathbf{V}_i(\mathbf{q},\mathbf{p})\mathrm{d}t + \sigma\mathbf{R}_i(\mathbf{q},t)\,, \end{aligned} \tag{2.37}$$

where $\mathbf{F}_i^{\mathrm{C}}(\mathbf{q})$ is the total conservative force acting on particle i

$$\mathbf{F}_i^{\mathrm{C}}(\mathbf{q}) = \sum_{j\neq i}\mathbf{F}_{ij}^{\mathrm{C}}(q_{ij}) = -\nabla_{\mathbf{q}_i}U(\mathbf{q})\,, \tag{2.38}$$

and $\mathbf{V}_i(\mathbf{q},\mathbf{p})$ and $\mathbf{R}_i(\mathbf{q},\mathbf{p},t)$ are defined respectively as

$$\mathbf{V}_i(\mathbf{q},\mathbf{p}) = \sum_{j\neq i}\omega^{\mathrm{D}}(q_{ij})(\hat{\mathbf{q}}_{ij}\cdot\mathbf{v}_{ij})\hat{\mathbf{q}}_{ij}\,, \tag{2.39}$$

$$\mathbf{R}_i(\mathbf{q},t) = \sum_{j\neq i}\omega^{\mathrm{R}}(q_{ij})\hat{\mathbf{q}}_{ij}\mathrm{d}\mathrm{W}_{ij}(t)\,, \tag{2.40}$$

where $\mathrm{d}\mathrm{W}_{ij}(t) = \mathrm{d}\mathrm{W}_{ji}(t)$ are independent increments of a Wiener process with mean zero and variance $\mathrm{d}t$ [65].

In fact, the DPD system (2.25) can be written in a more compact (vector) form:

$$\begin{aligned} \mathrm{d}\mathbf{q} &= \mathbf{M}^{-1}\mathbf{p}\mathrm{d}t\,, \\ \mathrm{d}\mathbf{p} &= -\nabla U(\mathbf{q})\mathrm{d}t - \gamma\boldsymbol{\Gamma}(\mathbf{q})\mathbf{M}^{-1}\mathbf{p}\mathrm{d}t + \sigma\boldsymbol{\Sigma}(\mathbf{q})\mathrm{d}\mathbf{W}\,, \end{aligned} \tag{2.41}$$

where $\mathbf{M}$ is the diagonal mass matrix, $\mathbf{W}$ is a vector of $S = dN(N-1)/2$ independent Wiener processes, and, projection matrices $\boldsymbol{\Gamma}(\mathbf{q}) \in \mathbb{R}^{dN\times dN}$ and $\boldsymbol{\Sigma}(\mathbf{q}) \in \mathbb{R}^{dN\times S}$ satisfy the following relation

$$\boldsymbol{\Gamma}(\mathbf{q}) = \boldsymbol{\Sigma}(\mathbf{q})\boldsymbol{\Sigma}(\mathbf{q})^T\,, \tag{2.42}$$

which can be thought of as the generalized fluctuation-dissipation theorem. In

particular, one can write down $\boldsymbol{\Gamma}(\mathbf{q})$ explicitly as

$$\boldsymbol{\Gamma}(\mathbf{q}) = \begin{pmatrix} \sum_{j\neq 1} \omega_{1j}^{\mathrm{D}} \mathbf{E}_{1j} & -\omega_{12}^{\mathrm{D}} \mathbf{E}_{12} & \cdots & -\omega_{1N}^{\mathrm{D}} \mathbf{E}_{1N} \\ -\omega_{21}^{\mathrm{D}} \mathbf{E}_{21} & \sum_{j\neq 2} \omega_{2j}^{\mathrm{D}} \mathbf{E}_{2j} & \cdots & -\omega_{2N}^{\mathrm{D}} \mathbf{E}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ -\omega_{N1}^{\mathrm{D}} \mathbf{E}_{N1} & -\omega_{N2}^{\mathrm{D}} \mathbf{E}_{N2} & \cdots & \sum_{j\neq N} \omega_{Nj}^{\mathrm{D}} \mathbf{E}_{Nj} \end{pmatrix}, \tag{2.43}$$

where $\omega_{ij}^{\mathrm{D}} = \omega^{\mathrm{D}}(q_{ij})$ is the weight function defined in the DPD system and $\mathbf{E}_{ij}$ is the d by d projection matrix on particles i and j:

$$\mathbf{E}_{ij} = \frac{(\mathbf{q}_i - \mathbf{q}_j)(\mathbf{q}_i - \mathbf{q}_j)^T}{\|\mathbf{q}_i - \mathbf{q}_j\|^2} = \hat{\mathbf{q}}_{ij} \hat{\mathbf{q}}_{ij}^T . \tag{2.44}$$

It should be noted that, when $\boldsymbol{\Sigma}(\mathbf{q})$ is an identity matrix, the DPD system (2.41) effectively reduces to the Langevin dynamics. Thus the latter can be viewed as a special case of the former.

2.6 Pairwise Nosé–Hoover–Langevin (PNHL) Thermostat

We have further generalized the Nosé–Hoover–Langevin (NHL) thermostat to obtain the pairwise Nosé–Hoover–Langevin (PNHL) thermostat [106], which is a momentum-conserving thermostat and thus applicable to the simulation of hydrodynamic behavior in complex fluids and polymers in mesoscales as in DPD.

The equations of motion of the PNHL system (for particle i) is given by

$$\begin{aligned} \mathrm{d}\mathbf{q}_i &= m_i^{-1} \mathbf{p}_i \mathrm{d}t \,, \\ \mathrm{d}\mathbf{p}_i &= \mathbf{F}_i^{\mathrm{C}}(\mathbf{q}) \mathrm{d}t - \xi \mathbf{V}_i(\mathbf{q}, \mathbf{p}) \mathrm{d}t \,, \\ \mathrm{d}\xi &= G(\mathbf{q}, \mathbf{p}) \mathrm{d}t - \tilde{\gamma} \xi \mathrm{d}t + \tilde{\sigma} \mathrm{dW} \,, \end{aligned} \tag{2.45}$$

where $G(\mathbf{q}, \mathbf{p})$ is the instantaneous accumulated deviation of the kinetic temperature away from the target temperature [6]:

$$G(\mathbf{q}, \mathbf{p}) = \mu^{-1} \sum_i \sum_{j>i} \omega^{\mathrm{D}}(q_{ij}) \left[(\mathbf{v}_{ij} \cdot \hat{\mathbf{q}}_{ij})^2 - k_{\mathrm{B}} T / m_{ij} \right] . \tag{2.46}$$

Coefficient constants $\tilde{\gamma}$ and $\tilde{\sigma}$ satisfy the fluctuation-dissipation theorem as in the Nosé–Hoover–Langevin thermostat (2.23). Thus, canonical ensemble is still preserved with the same invariant distribution (2.22) as in Nosé–Hoover and

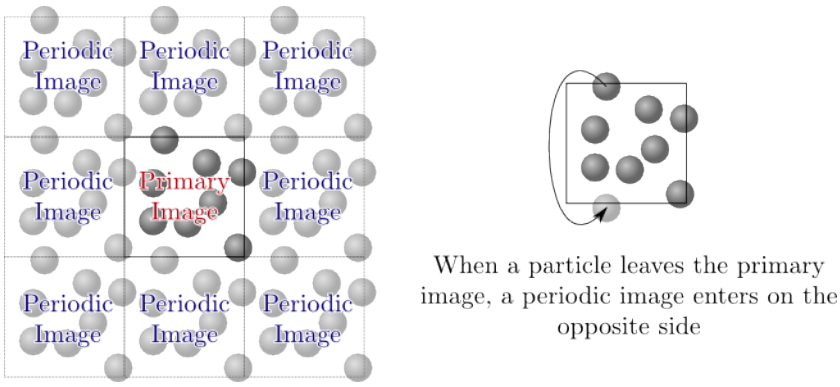


Figure 2.1: A two-dimensional illustration of the periodic boundary conditions (left) and the effective wraparound of particle positions (right).

Nosé–Hoover–Langevin:

$$\hat{\rho}_\beta(\mathbf{q},\mathbf{p},\xi) = \frac{1}{Z}\exp\left(-\beta H(\mathbf{q},\mathbf{p})\right)\exp\left(-\frac{\beta\mu\xi^2}{2}\right). \tag{2.47}$$

The PNHL thermostat (2.45) can also be written in a more compact form

$$\begin{aligned} \mathrm{d}\mathbf{q} &= \mathbf{M}^{-1}\mathbf{p}\mathrm{d}t\,, \\ \mathrm{d}\mathbf{p} &= -\nabla U(\mathbf{q})\mathrm{d}t - \xi\boldsymbol{\Gamma}(\mathbf{q})\mathbf{M}^{-1}\mathbf{p}\mathrm{d}t\,, \\ \mathrm{d}\xi &= G(\mathbf{q},\mathbf{p})\mathrm{d}t - \tilde{\gamma}\xi\mathrm{d}t + \tilde{\sigma}\mathrm{d}W\,, \end{aligned} \tag{2.48}$$

where $G(\mathbf{q},\mathbf{p})$ now can be written as

$$G(\mathbf{q},\mathbf{p}) = \mu^{-1}\left[\left(\mathbf{M}^{-1}\mathbf{p}\right)^T\boldsymbol{\Gamma}(\mathbf{q})\left(\mathbf{M}^{-1}\mathbf{p}\right) - \beta^{-1}\mathrm{Tr}\left(\boldsymbol{\Gamma}(\mathbf{q})\right)\right]. \tag{2.49}$$

2.7 Boundary Conditions

Computer simulations are often performed on a relatively small scale (i.e., a small number of particles in a relative sense, $\approx 10^5$ particles), which is limited by the available computing facilities and total computational cost in terms of execution time (the time taken for a double loop used to calculate the forces or potential energy without special techniques could be of order N^2). However, for the simulation of a bulk system (without suitable boundary conditions), a large fraction of particles would appear to lie on the cube surface, resulting in unphysical behavior [7]. In this section, we introduce two widely used boundary conditions that overcome the surface effects in molecular simulations.

2.7.1 Periodic Boundary Conditions

One of the most popular boundary conditions in equilibrium molecular dynamics is the periodic boundary conditions (PBC) [19]. In PBC, a primary cubic box, centered at the origin with lengths L_x, L_y, and L_z, in three dimensions, is replicated throughout space to form an infinite lattice. During the simulation, as a particle moves in the primary box, its periodic image moves in exactly the same way in each of the neighboring boxes. Therefore, once a particle leaves the primary box, one of its images will enter through the opposite face in a symmetric fashion with the same velocity. In this way, not only is the number of particles kept constant, the particles do not feel the walls at the boundaries any more, thus overcoming the surface effects. A two-dimensional schematic illustration of the PBC is given in Figure 2.1.

Denote the position and velocity of particle i as $\mathbf{q}_i = (q_i^x, q_i^y, q_i^z)$ and $\mathbf{v}_i = (v_i^x, v_i^y, v_i^z)$, respectively. PBC consists of two steps in practical implementation:

(a) Periodic Boundaries: when a particle leaves its primary box by crossing one of the boundaries, one of its images will enter through to the opposite face.

$$q_i^x \leftarrow q_i^x - L_x \cdot \text{round}(q_i^x / L_x) \tag{2.50}$$
$$q_i^y \leftarrow q_i^y - L_y \cdot \text{round}(q_i^y / L_y) \tag{2.51}$$
$$q_i^z \leftarrow q_i^z - L_z \cdot \text{round}(q_i^z / L_z) \tag{2.52}$$

where function "round($\cdot$)" returns the nearest integer to its argument.

(b) Minimum Image Convention: in order to correctly capture the transport properties of the system, particles are allowed to move on (rather than going back to the primary box when crossing the boundary), but interactions are evaluated with their nearest images.

$$q_{ij}^x \leftarrow q_{ij}^x - L_x \cdot \text{round}(q_{ij}^x / L_x) \tag{2.53}$$
$$q_{ij}^y \leftarrow q_{ij}^y - L_y \cdot \text{round}(q_{ij}^y / L_y) \tag{2.54}$$
$$q_{ij}^z \leftarrow q_{ij}^z - L_z \cdot \text{round}(q_{ij}^z / L_z) \tag{2.55}$$

where $q_{ij} = q_i - q_j$ is the distance between particle i and particle j.

It should be noted that there is no need to modify the velocity in PBC.

2.7.2 Lees–Edwards Boundary Conditions

As discussed in Section 1.2, nonequilibrium methods provide more efficient ways than equilibrium autocorrelation functions to extract transport coefficients (e.g., rheological properties such as shear stress and shear viscosity) from fluid dynamics simulations. Thus, there has been a rapidly growing interest in nonequilibri-

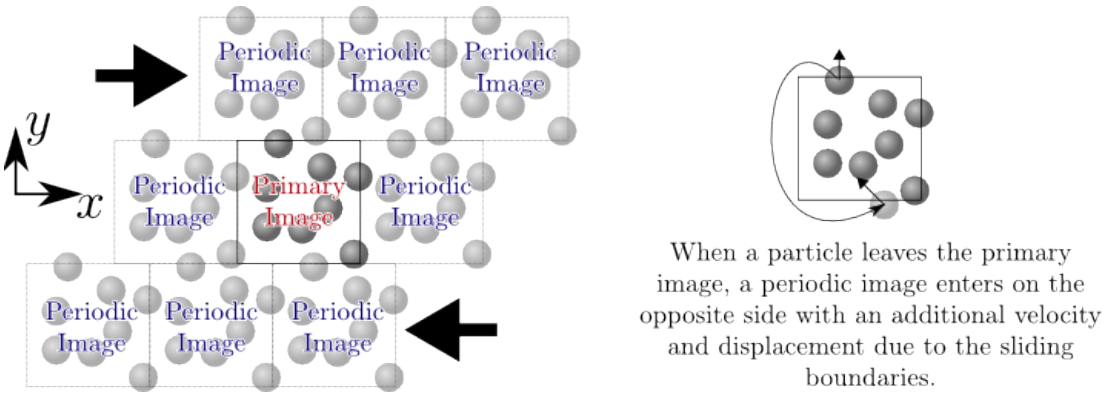


Figure 2.2: A two-dimensional illustration of the Lees–Edward sliding-brick boundary conditions.

um molecular dynamics (NEMD) simulations [51]. For instance, planar Couette flow, where a simple and steady shear flow is generated, is commonly employed as a numerical "viscometer" in particle-based methods to obtain transport coefficients [7, 51]. Furthermore, due to the fact that it is most relevant to real-life phenomena as well as its simplicity, planar Couette flow has also been widely adopted in real experiments.

In order to generate a simple shear flow in nonequilibrium molecular simulations, the periodic boundary conditions (PBC) have to be modified. A common way to achieve that is to apply the well-known Lees–Edwards boundary conditions (LEBC) [94]. In LEBC, the primary box remains centered at the origin, however, a uniform shear velocity profile is expected [51]

$$\mathbf{u} = \dot{\gamma} y \mathbf{e}^x \,, \tag{2.56}$$

where $\mathbf{e}^x$ is the unit vector in the x-direction and $\dot{\gamma}$ is the shear rate defined as

$$\dot{\gamma} = \frac{\mathrm{d}v^x}{\mathrm{d}y} \,. \tag{2.57}$$

Because of that, the LEBC is also called the "sliding brick" boundary conditions as shown in Figure 2.2.

Special attention has to be paid in LEBC when a particle is crossing the boundary in the y-direction. In this case, one of the images of the crossing particle will enter through the opposite face, but with both position and velocity modified in a proper way because of the streaming velocity (2.56).

The periodic boundary crossing is now handled as follows [7]:

$$N_{\mathrm{L}} = \mathrm{round}(q_i^y / L_y) \tag{2.58}$$
$$q_i^x \leftarrow q_i^x - N_{\mathrm{L}} \Delta q^x \tag{2.59}$$
$$q_i^x \leftarrow q_i^x - L_x \cdot \mathrm{round}(q_i^x / L_x) \tag{2.60}$$
$$q_i^y \leftarrow q_i^y - L_y N_{\mathrm{L}} \tag{2.61}$$
$$q_i^z \leftarrow q_i^z - L_z \cdot \mathrm{round}(q_i^z / L_z) \tag{2.62}$$
$$v_i^x \leftarrow v_i^x - N_{\mathrm{L}} \dot{\gamma} L_y \tag{2.63}$$

where N_{L} is the "rounded" number of layers (boxes) between the current position of particle i in the y-direction and the origin, and Δq^x is the displacement of the upper layer in Figure 2.2 during the elapsed time t from an appropriate origin, i.e.,

$$\Delta q^x = \dot{\gamma} L_y t \,. \tag{2.64}$$

The minimum image convention should now proceed as follows [7]:

$$N_{\mathrm{L}} = \mathrm{round}(q_{ij}^y / L_y) \tag{2.65}$$
$$q_{ij}^x \leftarrow q_{ij}^x - N_{\mathrm{L}} \Delta q^x \tag{2.66}$$
$$q_{ij}^x \leftarrow q_{ij}^x - L_x \cdot \mathrm{round}(q_{ij}^x / L_x) \tag{2.67}$$
$$q_{ij}^y \leftarrow q_{ij}^y - L_y N_{\mathrm{L}} \tag{2.68}$$
$$q_{ij}^z \leftarrow q_{ij}^z - L_z \cdot \mathrm{round}(q_{ij}^z / L_z) \tag{2.69}$$

Note that when the shear rate is zero, i.e., $\dot{\gamma} = 0$, the LEBC reduce to exactly the same as PBC.

Chapter 3

Numerical Methods

In this chapter, we investigate the numerical treatment of a wide range of methods introduced in the previous chapter. We systematically study the order of convergence to the invariant measure and practical performance of DPD and related momentum-conserving systems. Sections 3.3–3.7 are largely based on a recent publication [106] with B. Leimkuhler, while Section 3.8, where a nonsymmetric splitting method of PNHL has been shown to have second order convergence to the invariant measure for certain observables, has not appeared anywhere else.

3.1 Numerical Methods for Hamiltonian Dynamics

Consider an initial value problem

$$\dot{\mathbf{z}} = \mathbf{f}(\mathbf{z}) \,, \quad \mathbf{z}(0) = \mathbf{z}_0 \,, \tag{3.1}$$

in a multidimensional space where $\mathbf{z} \in \mathbb{R}^m$. Under suitable assumptions one may define a mapping from a point in phase space to the point t units later along the time-evolution starting from the initial point [101]. This map is referred to as the *flow map* and denoted by $\mathcal{F}_t$. In fact, $\mathcal{F}_t(\mathbf{z}_0) = \mathbf{z}(t)$ solves the initial value problem (3.1). However, in many cases, the system is not exactly solvable if $\mathbf{f}(\mathbf{z})$ is of a complicated (nonlinear) form, one may prefer to split up the vector field of the system into pieces (i.e., $\mathbf{f}(\mathbf{z}) = \mathbf{f}_1(\mathbf{z}) + \mathbf{f}_2(\mathbf{z}) + \dots$), and then solve each subsystem separately in certain order to approximate the true propagation in the phase space (see more discussions in [66, 101, 105]). Note that sometimes one can solve each subsystem exactly. For instance, a common splitting in separable

Hamiltonian dynamics (2.1)–(2.2), where $\mathbf{z} = [\mathbf{q}; \mathbf{p}]$, is as follows:

$$\mathrm{d}\begin{bmatrix} \mathbf{q} \\ \mathbf{p} \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{M}^{-1}\mathbf{p} \\ \mathbf{0} \end{bmatrix} \mathrm{d}t}_{\mathrm{A}} + \underbrace{\begin{bmatrix} \mathbf{0} \\ -\nabla U(\mathbf{q}) \end{bmatrix} \mathrm{d}t}_{\mathrm{B}}, \tag{3.2}$$

where one can easily solve each part "exactly" to generate exact flows, $\mathcal{F}_t^{\mathrm{A}}$ and $\mathcal{F}_t^{\mathrm{B}}$. That is, it can be easily seen that $\mathbf{p}$ is fixed in the A part, leading to exact solution to $\mathrm{d}\mathbf{q} = \mathbf{M}^{-1}\mathbf{p}\mathrm{d}t$, and similarly the B part is exactly solvable since $\mathbf{q}$ is constant. Then one can construct various numerical methods to solve the system.

The easiest choice might be an exact solver for one piece with stepsize h followed by another step for the other piece, which yields

$$\Phi_h^{\mathrm{BA}} = \mathcal{F}_h^{\mathrm{A}} \circ \mathcal{F}_h^{\mathrm{B}}, \tag{3.3}$$
$$\Phi_h^{\mathrm{AB}} = \mathcal{F}_h^{\mathrm{B}} \circ \mathcal{F}_h^{\mathrm{A}}. \tag{3.4}$$

The first method (3.3) is the symplectic Euler method, which reads

$$\mathbf{p}_{n+1} = \mathbf{p}_n - h\nabla U(\mathbf{q}_n), \tag{3.5}$$
$$\mathbf{q}_{n+1} = \mathbf{q}_n + h\mathbf{M}^{-1}\mathbf{p}_{n+1}, \tag{3.6}$$

while the second (3.4) is its adjoint method. Both methods are often referred to as the *Lie–Trotter splitting* [168]. It can be easily demonstrated that both methods are first order. However, one can also easily construct second order methods by integrating the splitting pieces in a symmetric fashion, which is known as the *Strang splitting* [161], such as

$$\Phi_h^{\mathrm{BAB}} = \mathcal{F}_{h/2}^{\mathrm{B}} \circ \mathcal{F}_h^{\mathrm{A}} \circ \mathcal{F}_{h/2}^{\mathrm{B}}, \tag{3.7}$$
$$\Phi_h^{\mathrm{ABA}} = \mathcal{F}_{h/2}^{\mathrm{A}} \circ \mathcal{F}_h^{\mathrm{B}} \circ \mathcal{F}_{h/2}^{\mathrm{A}}. \tag{3.8}$$

In this case, (3.7) is called "velocity Verlet" method, which reads

$$\mathbf{p}_{n+1/2} = \mathbf{p}_n - (h/2)\nabla U(\mathbf{q}_n), \tag{3.9}$$
$$\mathbf{q}_{n+1} = \mathbf{q}_n + h\mathbf{M}^{-1}\mathbf{p}_{n+1/2}, \tag{3.10}$$
$$\mathbf{p}_{n+1} = \mathbf{p}_{n+1/2} - (h/2)\nabla U(\mathbf{q}_{n+1}), \tag{3.11}$$

while (3.8) is the "position Verlet" method. It is to be understood that the steplengths associated with various operations are uniform and span the interval h. Thus the B step in the velocity Verlet (BAB) method (3.7) is taken with a step length of $h/2$, while A with a steplength of h. It can be shown that both Verlet methods are symplectic so that energy is well preserved during the simulation (see more details in [101]). Thus, Verlet methods have been widely used in molecular dynamics applications. Furthermore, arbitrary higher order methods can be constructed, such as the Yoshida (fourth order) method [177]. However, unlike the Verlet methods, those methods often require multiple computations of the force (the most expensive part in molecular simulations) per step, thus

significantly reducing their usefulness in practice.

3.2 Numerical Methods for Brownian and Langevin Dynamics

A systematic approach to thermodynamic bias in numerical computations has recently been proposed in studying the accuracy and efficiency of methods for Langevin dynamics [22, 99, 100]. The approach suggested is to determine the order of accuracy of a stochastic scheme in relation to its effective invariant measure (i.e., with respect to steady state averages computed using numerical trajectories). This technique has led to greatly improved numerical methods for Langevin dynamics as demonstrated in [99,100]. For instance, one may split the vector field of the Langevin dynamics (2.15)–(2.16) into three pieces, "A", "B", and "O":

$$\mathrm{d}\begin{bmatrix}\mathbf{q}\\ \mathbf{p}\end{bmatrix} = \underbrace{\begin{bmatrix}\mathbf{M}^{-1}\mathbf{p}\\ \mathbf{0}\end{bmatrix}\mathrm{d}t}_{\mathrm{A}} + \underbrace{\begin{bmatrix}\mathbf{0}\\ -\nabla U(\mathbf{q})\end{bmatrix}\mathrm{d}t}_{\mathrm{B}} + \underbrace{\begin{bmatrix}\mathbf{0}\\ -\gamma\mathbf{p}\mathrm{d}t + \sigma\mathbf{M}^{1/2}\mathrm{d}\mathbf{W}\end{bmatrix}}_{\mathrm{O}}, \quad (3.12)$$

where, as the fluctuation-dissipation theorem in DPD (2.31), σ is a constant coefficient satisfying the following relation with the friction constant γ:

$$\sigma^2 = 2\gamma k_{\mathrm{B}}T\,. \quad (3.13)$$

As in Hamiltonian dynamics, one can respectively solve the "A" and "B" pieces exactly. Furthermore, it is possible to derive the exact solution to the Ornstein–Uhlenbeck ("O") part,

$$\mathrm{d}\mathbf{p} = -\gamma\mathbf{p}\mathrm{d}t + \sigma\mathbf{M}^{1/2}\mathrm{d}\mathbf{W}\,. \quad (3.14)$$

Multiplying both sides of (3.14) by $\exp(\gamma t)$ and utilizing the fact that

$$\mathrm{d}(\exp(\gamma t)\mathbf{p}) = \exp(\gamma t)(\mathrm{d}\mathbf{p} + \gamma\mathbf{p}\mathrm{d}t) \quad (3.15)$$

yields

$$\mathrm{d}(\exp(\gamma t)\mathbf{p}) = \exp(\gamma t)\sigma\mathbf{M}^{1/2}\mathrm{d}\mathbf{W}\,. \quad (3.16)$$

Then integrating both sides from 0 to t gives

$$\exp(\gamma t)\mathbf{p}(t) = \mathbf{p}(0) + \int_0^t \exp(\gamma s)\sigma\mathbf{M}^{1/2}\,\mathrm{d}\mathbf{W}(s)\,, \quad (3.17)$$

which leads to the solution

$$\mathbf{p}(t) = \exp(-\gamma t)\mathbf{p}(0) + \exp(-\gamma t)\int_0^t \exp(\gamma s)\sigma\mathbf{M}^{1/2}\,\mathrm{d}\mathbf{W}(s)\,. \quad (3.18)$$

Denoting $\mathbf{Y}(t) = \int_0^t \exp(\gamma s)\, \mathrm{d}\mathbf{W}(s)$ with $\mathrm{d}\mathbf{W}(s)$ being a Wiener process, it is well known [90] that $\mathbf{Y}(t)$ is a Gaussian random variable with mean zero and variance (by using the Itō isometry: $\mathbb{E}[\int_0^t \mathbf{X}(s)\, \mathrm{d}\mathbf{W}(s)]^2 = \mathbb{E}\int_0^t [\mathbf{X}(s)]^2\, \mathrm{d}s$)

$$\mathrm{Var}\mathbf{Y}(t) = \int_0^t \exp(2\gamma s)\, \mathrm{d}s = \frac{e^{2\gamma t} - 1}{2\gamma}\,. \tag{3.19}$$

Therefore, the solution (3.18) with equivalent distribution is

$$\mathbf{p}(t) = e^{-\gamma t}\mathbf{p}(0) + \sigma\sqrt{\frac{1 - e^{-2\gamma t}}{2\gamma}}\mathbf{M}^{1/2}\mathbf{R}\,, \tag{3.20}$$

where $\mathbf{R}$ is a vector of independent and identically distributed (i.i.d.) standard normal random variables. In fact, the solution could be further generalized, incorporating the "B" part (which is essentially a constant when updating the momenta) as in the Stochastic Position Verlet method [117, 157]

$$\mathbf{p}(t) = \frac{\mathbf{F}(\mathbf{q})}{\gamma} + e^{-\gamma t}\left(\mathbf{p}(0) - \frac{\mathbf{F}(\mathbf{q})}{\gamma}\right) + \sigma\sqrt{\frac{1 - e^{-2\gamma t}}{2\gamma}}\mathbf{M}^{1/2}\mathbf{R}\,. \tag{3.21}$$

A number of numerical methods [25, 26, 48, 49, 117, 157, 170] for Langevin dynamics have been systematically studied and compared in various applications in [99, 100]. It turns out that different splitting methods give rather different performance in practice. Particularly, the so-called "BAOAB" method appears to be by far the best in terms of accuracy and efficiency for configurational sampling [99, 100]. The BAOAB method relies on exact solver for each piece and can be written as

$$\Phi_h^{\mathrm{BAOAB}} = \mathcal{F}_{h/2}^{\mathrm{B}} \circ \mathcal{F}_{h/2}^{\mathrm{A}} \circ \mathcal{F}_{h}^{\mathrm{O}} \circ \mathcal{F}_{h/2}^{\mathrm{A}} \circ \mathcal{F}_{h/2}^{\mathrm{B}}\,, \tag{3.22}$$

which reads

$$\begin{aligned}
\mathbf{p}_{n+1/3} &= \mathbf{p}_n - (h/2)\nabla U(\mathbf{q}_n)\,,\\
\mathbf{q}_{n+1/2} &= \mathbf{q}_n + (h/2)\mathbf{M}^{-1}\mathbf{p}_{n+1/3}\,,\\
\mathbf{p}_{n+2/3} &= e^{-\gamma h}\mathbf{p}_{n+1/3} + \sigma\sqrt{(1 - e^{-2\gamma h})/(2\gamma)}\mathbf{M}^{1/2}\mathbf{R}_n\,,\\
\mathbf{q}_{n+1} &= \mathbf{q}_{n+1/2} + (h/2)\mathbf{M}^{-1}\mathbf{p}_{n+2/3}\,,\\
\mathbf{p}_{n+1} &= \mathbf{p}_{n+1/2} - (h/2)\nabla U(\mathbf{q}_{n+1})\,.
\end{aligned}$$

It should be noted that only one force calculation is required at each step for BAOAB (i.e., the force computed at the end of each step will be reused at the beginning of the next step), the same as for alternative schemes.

It has been demonstrated that the BAOAB method has fourth order convergence to the invariant measure for configurational quantities in the high friction ($\gamma \to \infty$) limit [99]. Inspired by this superconvergence property, a second order discretization method for Brownian dynamics, the Leimkuhler–Matthews

method, was proposed [99, 102, 103]

$$\mathbf{q}_{n+1} = \mathbf{q}_n - h\nabla U(\mathbf{q}_n) + \sqrt{2h\beta^{-1}} \left(\frac{\mathbf{R}_n + \mathbf{R}_{n+1}}{2} \right) , \tag{3.23}$$

which is indeed very similar to the popular first order Euler–Maruyama method

$$\mathbf{q}_{n+1} = \mathbf{q}_n - h\nabla U(\mathbf{q}_n) + \sqrt{2h\beta^{-1}}\mathbf{R}_n , \tag{3.24}$$

except replacing the random realization by the average of two successive realizations. Both Euler–Maruyama and Leimkuhler–Matthews methods only require one force computation at each step, but rather surprisingly the latter provides one more order of convergence than the former essentially for "free". The stochastic Heun's method [90, 119] is second order and generally gives better accuracy in simulation than the Euler–Maruyama method (3.24), but it requires two force calculations at each step. It has been demonstrated that the newly proposed Leimkuhler–Matthews method (3.23) gives even more accurate long-time integration results in a number of applications than the stochastic Heun's method [103]. This again highlights the importance of optimal design of numerical methods.

In what follows we will systematically study the numerical treatment of momentum-conserving thermostats.

3.3 Numerical Methods for Dissipative Particle Dynamics

Due to the soft repulsive potential (2.33), the major advantage of the DPD method is that the stepsize used in simulations may be much larger than those of conventional MD simulations with Lennard-Jones internuclear potentials for instance. This feature is crucial especially when a very long simulation time is required. However, large stepsizes may result in errors in computed thermodynamic quantities. Whereas for molecular dynamics, simulations are performed at or near the stability threshold defined by stiff components such as harmonic bonds [100], DPD simulations may be perfectly stable over a wide range of stepsizes for which errors in averages are very large. There have been many attempts to develop accurate and efficient numerical methods that allow larger stepsizes. This is currently an active field of research.

In the early days of DPD, a number of integration schemes were proposed based on the well-recognized velocity Verlet scheme [172] widely used in classical MD simulations [7, 58]. Specific examples are the integrator of Groot and Warren [65] (GW), the method of Gibson et al. [60] (GCC), and the DPD velocity-Verlet integrator, which refers to as DPD-VV, of Besold et al. [16] (see more discussions in [171]). Both the GW and GCC integrators incorporate a parameter λ, which has to be chosen carefully for specific model parameters, to reduce

unphysical artifacts. Nevertheless, as reported in [16, 171], all the integrators mentioned above display pronounced artifacts, especially when the stepsize is large, due to the fact that the velocities and dissipative forces depend on each other implicitly and thus need to be updated in a self-consistent fashion.

In the spirit of the self-consistent leap-frog integrator introduced by Pagonabarraga et al. [129], Besold et al. [16] proposed a self-consistent velocity Verlet scheme, which we label SC-VV, in which, at the end of each iteration step, the velocity is "corrected" based on the newly calculated dissipative force until the deviation between the instantaneous kinetic temperature and the target temperature is less than a certain value of "tolerance". It should be noted that there is no such "correction" in the DPD-VV scheme but still the dissipative force is recalculated once, using the up-to-date velocities (momenta). A variant of the SC-VV scheme, SC-Th, introduced in [16], couples the original DPD system to an auxiliary Nosé–Hoover thermostat [75, 76, 127, 128] to provide direct kinetic temperature control. Overall, the self-consistent schemes do reduce unphysical artifacts to some extent, however, it is also well documented [31, 126] that they can be substantially slower than standard methods, depending how small the tolerance is. Therefore, computationally expensive self-consistent methods are not the choices present in typical software packages.

Although it has been demonstrated that relatively small stepsizes must be used in the DPD-VV scheme to produce correct static and dynamical properties [126, 171], the DPD-VV scheme remains one of the most popular methods for the DPD system in software packages due to its efficiency and ease of implementation (particularly in parallel computing). For this reason, the DPD-VV method has been chosen as the "benchmark" to compare with other schemes in this thesis.

Several novel integration schemes have been proposed over the years, such as the approach by den Otter and Clarke [43], the extended DPD method by Cotter and Reich [35], the multiple time step schemes by Jakobsen et al. [80], the Trotter-splitting methods by Thalmann and Farago [166], and most recently the algorithm by Goga et al. [62]. In this thesis, we focus on schemes that have been included in popular software packages (i.e., DPD-VV [16], the Peters thermostat [136], the Lowe–Andersen thermostat [114], and the Nosé–Hoover–Lowe–Andersen thermostat [160] in the DL_MESO [152] package) with two promising splitting methods by Shardlow [155] and De Fabritiis et al. [37], respectively. Particularly, Nikunen et al. [126] in 2003 showed that the performance of the Lowe–Andersen thermostat and Shardlow's scheme are superior to those of several other schemes for a number of different observables. Recently, Shardlow-like splitting algorithms have been further applied in DPD with various fixed conditions [112], and its parallel implementation has also been developed in [93].

The full details of the integration steps of each method described in this section are presented in a common language in Appendix A.

3.3.1 DPD Velocity-Verlet

For integration stepsize h, the simple DPD-VV integrator [16] reads

$$\mathbf{p}_i^{n+1/2} = \mathbf{p}_i^n + \left(h\mathbf{F}_i^{\mathrm{C}}(\mathbf{q}^n) + h\mathbf{F}_i^{\mathrm{D}}(\mathbf{q}^n, \mathbf{p}^n) + \sqrt{h}\mathbf{F}_i^{\mathrm{R}}(\mathbf{q}^n)\right)/2\,, \tag{3.25}$$

$$\mathbf{q}_i^{n+1} = \mathbf{q}_i^n + hm_i^{-1}\mathbf{p}_i^{n+1/2}\,, \tag{3.26}$$

$$\mathbf{p}_i^{n+1} = \mathbf{p}_i^{n+1/2} + \left(h\mathbf{F}_i^{\mathrm{C}}(\mathbf{q}^{n+1}) + h\mathbf{F}_i^{\mathrm{D}}(\mathbf{q}^{n+1}, \mathbf{p}^{n+1/2}) + \sqrt{h}\mathbf{F}_i^{\mathrm{R}}(\mathbf{q}^{n+1})\right)/2\,. \tag{3.27}$$

In fact, the DPD-VV scheme has two differences compared to the standard velocity Verlet method [172]: (1) the forces are not just the conventional conservative forces, but include dissipative and random forces, as well; (2) the dissipative forces have to be updated for a second time at the end of each integration step by using the up-to-date velocities (momenta), $\mathbf{F}_i^{\mathrm{D}}(\mathbf{q}^{n+1}, \mathbf{p}^{n+1})$, with the first update taking place right after the positions are updated at each step (see more details in Appendix A). It has been shown that the performance of the DPD-VV method would be significantly improved [16, 171] by simply doing the additional update of the dissipative forces in each step, which is actually not time-consuming if one makes use of computation-saving devices such as Verlet neighbor lists [172]. Note that both the GW integrator [65] of Groot and Warren and the modified Verlet method mentioned by Shardlow in [155] do not incorporate the additional update.

It is important to observe that, unlike the standard velocity Verlet method (second order), the DPD-VV scheme should only give first order convergence for the invariant measure (see more details in Section 3.6) due to the fact that the momentum is not updated in a symmetric manner. However, second order convergence for averages was observed in the numerical experiments in Section 3.7. Moreover, the term $\sqrt{h}/2$ multiplying the random forces, which would be expected to be different in this type of splitting of Langevin dynamics, must be present when random forces are reused in the subsequent step; it ensures that the diffusion coefficient of the particles is independent of the integration timestep (see [65] for further discussion).

3.3.2 Shardlow's Splitting Method

Splitting techniques were studied by Shardlow [155] based on dividing the vector field of the DPD system into three parts, the first two of which represent the vector field of the Hamiltonian system associated with kinetic and potential energies, and the last term is the remaining Langevin equation (actually Ornstein–Uhlenbeck process with positions fixed) involving the dissipative and random forces. Two integrators, termed there S1 and S2, have been proposed for treating this system in [155]. Only the S1 method will be examined here as it is more efficient than S2. This scheme relies on the method of Brünger, Brooks, and Karplus (BBK) [25] to solve the Langevin part, followed by the standard velocity Verlet scheme for

the conservative part.

In describing Shardlow's method or another splitting scheme, we use the formal notation of the generator of the diffusion as in, for example, [37, 153, 166].

We first separate the system of stochastic differential equations for DPD (2.37) into three pieces, which we label as A, B, and O:

$$\mathrm{d}\begin{bmatrix} \mathbf{q}_i \\ \mathbf{p}_i \end{bmatrix} = \underbrace{\begin{bmatrix} m_i^{-1}\mathbf{p}_i \\ \mathbf{0} \end{bmatrix}\mathrm{d}t}_{\mathrm{A}} + \underbrace{\begin{bmatrix} \mathbf{0} \\ \mathbf{F}_i^{\mathrm{C}} \end{bmatrix}\mathrm{d}t}_{\mathrm{B}} + \underbrace{\begin{bmatrix} \mathbf{0} \\ -\gamma\mathbf{V}_i\mathrm{d}t + \sigma\mathbf{R}_i \end{bmatrix}}_{\mathrm{O}} . \tag{3.28}$$

The generators for each part of the SDE may be written out as follows:

$$\mathcal{L}_{\mathrm{A}} = \sum_i \frac{\mathbf{p}_i}{m_i} \cdot \nabla_{\mathbf{q}_i} , \tag{3.29}$$

$$\mathcal{L}_{\mathrm{B}} = \sum_i \mathbf{F}_i^{\mathrm{C}} \cdot \nabla_{\mathbf{p}_i} = -\sum_i \nabla_{\mathbf{q}_i} U(\mathbf{q}) \cdot \nabla_{\mathbf{p}_i} , \tag{3.30}$$

$$\mathcal{L}_{\mathrm{O}} = \sum_i \sum_{j \neq i} \left(-\gamma\omega^{\mathrm{D}}(q_{ij})(\hat{\mathbf{q}}_{ij} \cdot \mathbf{v}_{ij}) + \frac{\sigma^2}{2}\left[\omega^{\mathrm{R}}(q_{ij})\right]^2 \hat{\mathbf{q}}_{ij} \cdot \left(\nabla_{\mathbf{p}_i} - \nabla_{\mathbf{p}_j}\right) \right) \hat{\mathbf{q}}_{ij} \cdot \nabla_{\mathbf{p}_i} . \tag{3.31}$$

Thus, the generator for the DPD system can be written as

$$\mathcal{L}_{\mathrm{DPD}} = \mathcal{L}_{\mathrm{A}} + \mathcal{L}_{\mathrm{B}} + \mathcal{L}_{\mathrm{O}} . \tag{3.32}$$

The flow map (or phase space propagator) of the system may be written in the shorthand notation

$$\mathcal{F}_t = e^{t\mathcal{L}_{\mathrm{DPD}}} , \tag{3.33}$$

where the exponential map is here used formally to denote the solution operator. Approximations of $\mathcal{F}_t$ may then be obtained as products (taken in different arrangements) of exponentials of the various terms of the splitting. For example, the phase space propagation of Shardlow's S1 splitting method [155], termed DPD-S1, can be written as

$$\exp\left(h\hat{\mathcal{L}}_{\mathrm{DPD-S1}}\right) = \exp\left(h\mathcal{L}_{\mathrm{O}}\right) \exp\left(\frac{h}{2}\mathcal{L}_{\mathrm{B}}\right) \exp\left(h\mathcal{L}_{\mathrm{A}}\right) \exp\left(\frac{h}{2}\mathcal{L}_{\mathrm{B}}\right) , \tag{3.34}$$

where h is the stepsize and $\exp\left(h\mathcal{L}_f\right)$ represents the phase space propagator associated with the corresponding vector field f. In Shardlow's approach, the vector field O is further split into each interacting pair. Therefore, the propagation of the O part is broken down into many terms:

$$\exp\left(h\hat{\mathcal{L}}_{\mathrm{O}}\right) = \exp\left(h\mathcal{L}_{\mathrm{O}_{N-1,N}}\right) \cdots \exp\left(h\mathcal{L}_{\mathrm{O}_{1,3}}\right) \exp\left(h\mathcal{L}_{\mathrm{O}_{1,2}}\right) ,$$

where the operator associated with each interacting pair is defined as

$$\mathcal{L}_{\mathrm{O}_{i,j}} = \left(-\gamma\omega^{\mathrm{D}}(q_{ij})(\hat{\mathbf{q}}_{ij}\cdot\mathbf{v}_{ij}) + \frac{\sigma^2}{2}\left[\omega^{\mathrm{R}}(q_{ij})\right]^2 \hat{\mathbf{q}}_{ij}\cdot\left(\nabla_{\mathbf{p}_i} - \nabla_{\mathbf{p}_j}\right)\right)\hat{\mathbf{q}}_{ij}\cdot\nabla_{\mathbf{p}_i}\,. \tag{3.35}$$

Each interacting pair preserves the invariant distribution ρ_β (2.34). As a shorthand, we may term the DPD-S1 method OBAB (similarly, the S2 method of Shardlow would be equivalent to OBABO in the same language).

3.3.3 DPD-Trotter Scheme

The Trotter formula [161] that has been widely used in molecular simulations was investigated and further applied to split the DPD generator (3.32) in an "optimal" way to reduce artifacts and maintain good temperature control [37, 153]. A new scheme, referred to as DPD-Trotter, was introduced but few numerical simulations have been presented and therefore we incorporate it in the comparisons.

In the stochastic DPD-Trotter scheme, the standard DPD system (2.37) is split into two parts, which are labeled "A" and "S" as indicated below

$$\mathrm{d}\begin{bmatrix}\mathbf{q}_i\\ \mathbf{p}_i\end{bmatrix} = \underbrace{\begin{bmatrix} m_i^{-1}\mathbf{p}_i\\ \mathbf{0}\end{bmatrix}\mathrm{d}t}_{\mathrm{A}} + \underbrace{\begin{bmatrix}\mathbf{0}\\ \mathbf{F}_i^{\mathrm{C}}\mathrm{d}t - \gamma\mathbf{V}_i\mathrm{d}t + \sigma\mathbf{R}_i\end{bmatrix}}_{\mathrm{S}}\,. \tag{3.36}$$

The corresponding operator of part A is exactly the same as in Shardlow's method, while the operator of part S is actually the sum of the operators of B and O defined above

$$\mathcal{L}_{\mathrm{S}} = \mathcal{L}_{\mathrm{B}} + \mathcal{L}_{\mathrm{O}}\,. \tag{3.37}$$

As in Shardlow's method, the vector field S is further split into each interacting pair; these pair interactions are exactly solvable (in the sense of distributional fidelity). In fact, for $j > i$, subtracting $\mathrm{d}\mathbf{v}_j$ from $\mathrm{d}\mathbf{v}_i$ and multiplying $\hat{\mathbf{q}}_{ij}$ on both sides gives

$$m_{ij}\mathrm{d}v_{ij} = F_{ij}^{\mathrm{C}}(q_{ij})\mathrm{d}t - \gamma\omega^{\mathrm{D}}(q_{ij})v_{ij}\mathrm{d}t + \sigma\omega^{\mathrm{R}}(q_{ij})\mathrm{d}\mathrm{W}_{ij}\,, \tag{3.38}$$

where $m_{ij} = m_i m_j/(m_i+m_j)$ is the "reduced mass", $v_{ij} = \hat{\mathbf{q}}_{ij}\cdot\mathbf{v}_{ij}$ and $F_{ij}^{\mathrm{C}}(q_{ij})$ is the magnitude of the conservative force (2.27). The above equation is an Ornstein–Uhlenbeck process with the exact (in the sense of distributions) solution [90]

$$v_{ij}(t) = \frac{F_{ij}^{\mathrm{C}}}{\tau m_{ij}} + e^{-\tau t}\left(v_{ij}(0) - \frac{F_{ij}^{\mathrm{C}}}{\tau m_{ij}}\right) + \sqrt{\frac{k_{\mathrm{B}}T(1-e^{-2\tau t})}{m_{ij}}}\mathrm{R}_{ij}\,, \tag{3.39}$$

where $\tau = \gamma\omega^{\mathrm{D}}/m_{ij}$, $v_{ij}(0)$ are the initial relative velocities and R_{ij} are normally distributed variables with zero mean and unit variance. Thus the increment

velocities can be obtained as

$$\Delta v_{ij} = v_{ij}(t) - v_{ij}(0) = \left(v_{ij}(0) - \frac{F_{ij}^{\mathrm{C}}}{\tau m_{ij}} \right) (e^{-\tau t} - 1) + \sqrt{\frac{k_{\mathrm{B}} T (1 - e^{-2\tau t})}{m_{ij}}} \mathrm{R}_{ij} \,, \tag{3.40}$$

and the corresponding momenta can be updated by

$$\mathbf{p}_i^{n+1} = \mathbf{p}_i^n + m_{ij} \Delta v_{ij} \hat{\mathbf{q}}_{ij}^n \,, \tag{3.41}$$

$$\mathbf{p}_j^{n+1} = \mathbf{p}_j^n - m_{ij} \Delta v_{ij} \hat{\mathbf{q}}_{ij}^n \,, \tag{3.42}$$

which defines the propagator $e^{h\mathcal{L}_{\mathrm{S}_{i,j}}}$ for each interacting pair. Overall, the propagator of the DPD-Trotter scheme can be written as

$$\exp\left(h\hat{\mathcal{L}}_{\mathrm{DPD-Trotter}}\right) = \exp\left(\frac{h}{2}\mathcal{L}_{\mathrm{S}}\right) \exp\left(h\mathcal{L}_{\mathrm{A}}\right) \exp\left(\frac{h}{2}\mathcal{L}_{\mathrm{S}}\right) \,,$$

where the momentum part is defined by

$$\exp\left(\frac{h}{2}\hat{\mathcal{L}}_{\mathrm{S}}\right) = \exp\left(\frac{h}{2}\mathcal{L}_{\mathrm{S}_{N-1,N}}\right) \cdots \exp\left(\frac{h}{2}\mathcal{L}_{\mathrm{S}_{1,3}}\right) \exp\left(\frac{h}{2}\mathcal{L}_{\mathrm{S}_{1,2}}\right) \,.$$

3.4 Numerical Methods for Alternative Methods

3.4.1 Lowe–Andersen Thermostat

The Schmidt number, S_{c}, which is the ratio of the kinematic viscosity ν to the diffusion coefficient D, is an important quantity that characterizes the dynamical behavior of fluids. In a typical fluid flow, water for example, momentum can be transported more rapidly than particles, resulting in Schmidt number of order 10^3. However, as it was reported in [65], the standard DPD system (2.37) produces a gas-like Schmidt number of order one. Depending on the application, this could be a significant disadvantage for simulating more fluid-like system, although recent work by Fan et al. [53] indicates that the Schmidt number of the standard DPD system can be varied by modifying the weight function and increasing the cutoff radius.

To overcome the issue of low Schmidt number in the standard DPD system, instead of using a Langevin thermostat to reequilibrate the system, Lowe [114] employed a pairwise stochastic momentum-conserving Andersen thermostat [8], in which after updating the positions and momenta due to the conservative forces only by using the standard velocity Verlet method, the momenta are updated, with probability $P = \Gamma h$, by reselecting the relative velocities for interacting pairs

from the Maxwell–Boltzmann distribution,

$$\mathbf{p}_i \leftarrow \mathbf{p}_i + \Delta\mathbf{p}_{ij}\,, \tag{3.43}$$
$$\mathbf{p}_j \leftarrow \mathbf{p}_j - \Delta\mathbf{p}_{ij}\,, \tag{3.44}$$

with

$$\Delta\mathbf{p}_{ij} = m_{ij}\left[\mathrm{R}_{ij}\sqrt{k_\mathrm{B}T/m_{ij}} - \hat{\mathbf{q}}_{ij}\cdot\mathbf{v}_{ij}\right]\hat{\mathbf{q}}_{ij}\,, \tag{3.45}$$

where R_{ij} are Gaussian random variables with zero mean and unit variance. The parameter Γ can be thought of as the stochastic randomization frequency with upper limit $1/h$. Lowe's method is frequently referred to as the Lowe–Andersen (LA) thermostat, which still conserves the momentum and hydrodynamics. The additional Andersen thermostat does not change the distribution of the system [8], therefore the same invariant distribution (2.34) as in the standard DPD system is maintained. Most important, the LA thermostat is capable of varying the Schmidt number by changing the parameter Γ. When the probability of further updating the momentum is high (large P), the viscosity is high and diffusion coefficient is low, resulting in large Schmidt number in the regime of typical fluids. The LA thermostat has been applied in molecular dynamics simulations by Koopman and Lowe [91]. It is worth mentioning that the generator of the LA thermostat does not converge to that of the standard DPD system as $h \to 0$.

3.4.2 Peters Thermostat

Based on various numerical studies on the DPD system, all the numerical methods based on discretization of the equation of motion are dependent on the stepsize. In order to reduce the dependence, Peters generalized the Lowe–Andersen (LA) thermostat and presented another approach to reequilibrate the system [136]. Following the update of the conservative part by using the standard velocity Verlet scheme, the momenta of all interacting pairs will be further updated (not in random order as in the original paper, which does not have much effect on the results but reduces computational cost) as follows

$$\mathbf{p}_i \leftarrow \mathbf{p}_i + \Delta\mathbf{p}_{ij}\,, \tag{3.46}$$
$$\mathbf{p}_j \leftarrow \mathbf{p}_j - \Delta\mathbf{p}_{ij}\,, \tag{3.47}$$

with

$$\Delta\mathbf{p}_{ij} = \left[-\gamma_{ij}(\hat{\mathbf{q}}_{ij}\cdot\mathbf{v}_{ij})h + \sigma_{ij}\sqrt{h}\mathrm{R}_{ij}\right]\hat{\mathbf{q}}_{ij}\,, \tag{3.48}$$

where R_{ij} are the standard Gaussian random variables as in the Lowe–Andersen thermostat, and the coefficients γ_{ij} and σ_{ij} satisfies the following condition

$$\sigma_{ij} = \sqrt{2k_\mathrm{B}T\gamma_{ij}\left[1 - \gamma_{ij}h/(2m_{ij})\right]}\,,$$

which reduces to the fluctuation-dissipation theorem (2.31) in standard DPD in the limit of $h \to 0$. Two possible choices of the coefficients was presented in the original paper [136], but only "Scheme II", which has less restriction on the choice of stepsize h, was chosen to compare with other methods in this thesis. In the $h \to 0$ limit, it can be easily shown that the generator of the Peters thermostat converges to that of the standard DPD system and therefore conserves the canonical ensemble. Unfortunately, numerical simulations in the original paper showed that the thermostat still exhibits significant deviation both in static (kinetic temperature) and dynamical (diffusion coefficient) quantities in standard model settings (model B in the language of [126]) when the timestep is above 0.05.

3.4.3 Nosé–Hoover–Lowe–Andersen Thermostat

Recently, Stoyanov and Groot combined the Lowe–Andersen (LA) thermostat with a Nosé–Hoover-like thermostat to construct a local Galilean invariant stochastic momentum-conserving thermostat, the Nosé–Hoover–Lowe–Andersen (NHLA) thermostat [160], to achieve direct kinetic temperature control. Unlike the LA thermostat, a modified version of the velocity Verlet scheme is used to update the positions and velocities at the start of each integration step

$$\mathbf{q}_i \leftarrow \mathbf{q}_i + h\mathbf{v}_i + h^2\mathbf{F}_i^{\mathrm{C}}(\mathbf{q})/2\,, \tag{3.49}$$

$$\mathbf{v}_i \leftarrow \mathbf{v}_i + h\mathbf{F}_i^{\mathrm{C}}(\mathbf{q})/2\,, \tag{3.50}$$

which is followed by the calculation of the updated conservative forces. After that, the fraction $(1 - P)$ of interacting pairs that do not have their relative velocities stochastically reselected are thermalized by a deterministic method instead. For each such pair, the dissipative force is calculated

$$\mathbf{F}_{ij}^{\mathrm{D}} = \alpha\omega^{\mathrm{R}}(q_{ij})(\hat{\mathbf{q}}_{ij} \cdot \mathbf{v}_{ij})\hat{\mathbf{q}}_{ij}\,,$$

where α is a coupling parameter chosen as $0.9/(\rho h)$ in this thesis such that, overall, the dissipative force defined above is the same as that in the original paper [160], and ρ is the particle density. The dissipative force of each particle is updated

$$\mathbf{F}_i^{\mathrm{D}} \leftarrow \mathbf{F}_i^{\mathrm{D}} + \mathbf{F}_{ij}^{\mathrm{D}}\,, \tag{3.51}$$

$$\mathbf{F}_j^{\mathrm{D}} \leftarrow \mathbf{F}_j^{\mathrm{D}} - \mathbf{F}_{ij}^{\mathrm{D}}\,. \tag{3.52}$$

Then, after the further update of the velocities

$$\mathbf{v}_i \leftarrow \mathbf{v}_i + h\mathbf{F}_i^{\mathrm{C}}(\mathbf{q})/2\,, \tag{3.53}$$

the momenta are corrected by

$$\mathbf{p}_i \leftarrow \mathbf{p}_i + h(1 - \tilde{T}_{\mathrm{k}}/T_0)\mathbf{F}_i^{\mathrm{D}}\,, \tag{3.54}$$

where T_0 is the desired temperature and the momentary kinetic temperature $\tilde{T}_\mathrm{k}$ is calculated from the relative velocities at the time of calculating the conservative force to enhance the computational efficiency and save memory space (this is slightly different from the approach in the original paper which uses the updated Verlet neighbor lists but the stored velocities from the previous integration step)

$$k_\mathrm{B}\tilde{T}_\mathrm{k} = \frac{\sum_{j\neq i}\omega(q_{ij})m_{ij}(\mathbf{v}_i-\mathbf{v}_j)\cdot(\mathbf{v}_i-\mathbf{v}_j)}{d\sum_{j\neq i}\omega(q_{ij})}\,, \tag{3.55}$$

where $\omega(q_{ij})$ is a smearing function chosen as

$$\omega(q_{ij}) = \begin{cases} 1\,, & q_{ij} < r_\mathrm{c}\,; \\ 0\,, & q_{ij} \geq r_\mathrm{c}\,. \end{cases} \tag{3.56}$$

Finally, the momenta are further updated as in the LA thermostat (Equations (3.43)–(3.44)).

The factor $(1-\tilde{T}_\mathrm{k}/T_0)$ in (3.54) acts like the "friction coefficient" to tune the kinetic temperature of the system. It is actually not a dynamical variable as in the standard Nosé–Hoover thermostat, instead is more closely related to the Berendsen thermostat [13]. As reported in [160], the NHLA thermostat maintains an order of magnitude improvement in kinetic temperature and can also vary the Schmidt number by several orders of magnitude as in the LA thermostat. However, with large stepsizes that maintain good kinetic temperature control (1% relative error) in the NHLA thermostat, substantial errors in configurational temperature were reported [5], which indicates that the system temperature was not sampled correctly. It is worthy of mention that a slightly modified integration strategy was used in [5], which does not have much effect on the results as discussed in the original paper [160]. Moreover, the generator of the NHLA thermostat does not converge to that of the standard DPD system as $h \to 0$, and, it has not been shown that the NHLA thermostat preserves the canonical ensemble.

3.5 Numerical Methods for Extended Variable Momentum-Conserving Thermostats

3.5.1 Pairwise Nosé–Hoover Thermostat

In all standard DPD methods (Section 3.3) and alternative methods (Section 3.4) in DPD simulations, a random number has to be generated for each interacting pair, which can be very time-consuming depending on the particle density and cutoff radius, and, becomes trickier when parallel computing techniques (multiple processors for domain-decomposed cells) are used: it requires additional, even

substantial, effort to communicate interacting particles in different cells [3, 137]. Based on the Nosé–Hoover–Lowe–Andersen (NHLA) thermostat by Stoyanov and Groot [160], Allen and Schmid [6] presented a new thermostat of the Nosé–Hoover type, in which stochastic terms were totally removed and the constant friction coefficient was replaced by a dynamical variable that was driven by the difference between the instantaneous kinetic temperature and the target temperature. The so-called pairwise Nosé–Hoover (PNH) thermostat conserves the momentum and is also Galilean-invariant, therefore correct hydrodynamics are still expected to be generated and it can be used in DPD simulations. Moreover, one may find it useful in nonequilibrium molecular dynamics (NEMD) to reduce unphysical behaviors (see more discussions in [6]).

The equations of motion of the PNH thermostat (for particle i) is given by

$$\begin{aligned} \mathrm{d}\mathbf{q}_i &= m_i^{-1}\mathbf{p}_i \mathrm{d}t\,, \\ \mathrm{d}\mathbf{p}_i &= \mathbf{F}_i^{\mathrm{C}}(\mathbf{q})\mathrm{d}t - \xi \mathbf{V}_i(\mathbf{q},\mathbf{p})\mathrm{d}t\,, \\ \mathrm{d}\xi &= G(\mathbf{q},\mathbf{p})\mathrm{d}t\,, \end{aligned} \tag{3.57}$$

where $\mathbf{V}_i(\mathbf{q},\mathbf{p})$ is defined in (2.39), ξ is the dynamical variable, and $G(\mathbf{q},\mathbf{p})$ is the instantaneous accumulated deviation of the kinetic temperature away from the target temperature [6]

$$G(\mathbf{q},\mathbf{p}) = \mu^{-1}\sum_i \sum_{j>i} \omega^{\mathrm{D}}(q_{ij}) \left[(\mathbf{v}_{ij}\cdot\hat{\mathbf{q}}_{ij})^2 - k_{\mathrm{B}}T/m_{ij} \right] . \tag{3.58}$$

Assuming ergodicity, canonical ensemble is still preserved with the same invariant distribution as that of PNHL (3.77).

A nonsymmetric integration algorithm (see Appendix A) was applied in the original paper [6] to solve the system:

$$\begin{aligned} \mathbf{p}_i^{n+1/2} &= \mathbf{p}_i^n + h\left(\mathbf{F}_i^{\mathrm{C}}(\mathbf{q}^n) - \xi^n \mathbf{V}_i(\mathbf{q}^n, \mathbf{p}^{n-1/2})\right)/2\,, \\ \xi^{n+1/2} &= \xi^n + hG(\mathbf{q}^n, \mathbf{p}^{n-1/2})/2, \\ \mathbf{q}_i^{n+1} &= \mathbf{q}_i^n + hm_i^{-1}\mathbf{p}_i^{n+1/2}\,, \\ \mathbf{p}_i^{n+1} &= \mathbf{p}_i^{n+1/2} + h\left(\mathbf{F}_i^{\mathrm{C}}(\mathbf{q}^{n+1}) - \xi^{n+1/2}\mathbf{V}_i(\mathbf{q}^{n+1}, \mathbf{p}^{n+1/2})\right)/2\,, \\ \xi^{n+1} &= \xi^{n+1/2} + hG(\mathbf{q}^{n+1}, \mathbf{p}^{n+1/2})/2\,. \end{aligned}$$

3.5.2 Pairwise Nosé–Hoover–Langevin Thermostat

In order to enhance the ergodicity and have a better temperature control, we have reformulated the pairwise Nosé–Hoover (PNH) thermostat (3.57) to form a pairwise Nosé–Hoover–Langevin (PNHL) thermostat (2.45) by adding a Langevin thermostat to the additional variable ξ in such a way that the invariant distribution (3.77) is not violated. As in the PNH thermostat, the PNHL thermostat has

the potential of being useful in NEMD, but we focus on the application of DPD in this thesis.

The vector field of the PNHL system can be split into five pieces below such that each piece can be solved "exactly",

$$\mathrm{d}\begin{bmatrix}\mathbf{q}_i\\ \mathbf{p}_i\\ \xi\end{bmatrix} = \underbrace{\begin{bmatrix} m_i^{-1}\mathbf{p}_i\\ \mathbf{0}\\ 0\end{bmatrix}\mathrm{d}t}_{\mathrm{A}} + \underbrace{\begin{bmatrix}\mathbf{0}\\ \mathbf{F}_i^{\mathrm{C}}\\ 0\end{bmatrix}\mathrm{d}t}_{\mathrm{B}} + \underbrace{\begin{bmatrix}\mathbf{0}\\ -\xi\mathbf{V}_i\\ 0\end{bmatrix}\mathrm{d}t}_{\mathrm{C}} + \underbrace{\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\\ G\end{bmatrix}\mathrm{d}t}_{\mathrm{D}} + \underbrace{\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\\ -\tilde{\gamma}\xi\mathrm{d}t + \tilde{\sigma}\mathrm{dW}\end{bmatrix}}_{\mathrm{O}}.$$

Note that the operators of A and B are exactly the same as defined in (3.29) and (3.30), respectively. We can also write down the operators for the remaining pieces as

$$\begin{aligned}\mathcal{L}_{\mathrm{C}} &= -\xi\sum_i \mathbf{V}_i(\mathbf{q},\mathbf{p})\cdot\nabla_{\mathbf{p}_i}\,,\\ \mathcal{L}_{\mathrm{D}} &= G(\mathbf{q},\mathbf{p})\frac{\partial}{\partial\xi}\,,\\ \mathcal{L}_{\mathrm{O}} &= -\tilde{\gamma}\xi\frac{\partial}{\partial\xi} + \frac{\tilde{\sigma}^2}{2}\frac{\partial^2}{\partial\xi^2}\,.\end{aligned}$$

The generator O here is slightly different from that in (3.31) which involves pairwise terms in DPD. Overall, the generator for the PNHL system can be written as

$$\mathcal{L}_{\mathrm{PNHL}} = \mathcal{L}_{\mathrm{A}} + \mathcal{L}_{\mathrm{B}} + \mathcal{L}_{\mathrm{C}} + \mathcal{L}_{\mathrm{D}} + \mathcal{L}_{\mathrm{O}}\,. \tag{3.59}$$

There are a variety of approaches to splitting this system. For example, we could use the same technique as in DPD-Trotter scheme (Section 3.3.3) to solve part C, but without the conservative and stochastic terms. Also, the O part may be solved exactly using the analytical weak solution of the Ornstein–Uhlenbeck process [90]

$$\xi(t) = e^{-\tilde{\gamma}t}\xi(0) + \sqrt{k_{\mathrm{B}}T(1-e^{-2\tilde{\gamma}t})/\mu}\mathrm{R}\,, \tag{3.60}$$

where $\xi(0)$ is the initial value of the additional variable and R are uncorrelated independent standard normal random variables.

Interestingly as noted in the setting of Langevin dynamics [99, 100], integrating those different splitting pieces in different orders gives dramatically different performance in terms of kinetic temperature control and other configurational quantities. We present here two approaches to integrate the PNHL system: first in a symmetric manner, termed PNHL-S, and the other nonsymmetric, termed PNHL-N. The propagators of the two schemes (see more details in Appendix A)

may be defined as

$$e^{h\hat{\mathcal{L}}_{\text{PNHL-S}}} = e^{\frac{h}{2}\mathcal{L}_{\text{A}}} e^{\frac{h}{2}\mathcal{L}_{\text{B}}} e^{\frac{h}{2}\mathcal{L}_{\text{C}}} e^{\frac{h}{2}\mathcal{L}_{\text{D}}} e^{h\mathcal{L}_{\text{O}}} e^{\frac{h}{2}\mathcal{L}_{\text{D}}} e^{\frac{h}{2}\mathcal{L}_{\text{C}}} e^{\frac{h}{2}\mathcal{L}_{\text{B}}} e^{\frac{h}{2}\mathcal{L}_{\text{A}}} , \tag{3.61}$$

and

$$e^{h\hat{\mathcal{L}}_{\text{PNHL-N}}} = e^{\frac{h}{2}\mathcal{L}_{\text{A}}} e^{\frac{h}{2}\mathcal{L}_{\text{B}}} e^{\frac{h}{2}\mathcal{L}_{\text{C}}} e^{\frac{h}{2}\mathcal{L}_{\text{D}}} e^{h\mathcal{L}_{\text{O}}} e^{\frac{h}{2}\mathcal{L}_{\text{D}}} e^{\frac{h}{2}\mathcal{L}_{\text{C}}} e^{\frac{h}{2}\mathcal{L}_{\text{A}}} e^{\frac{h}{2}\mathcal{L}_{\text{B}}} . \tag{3.62}$$

It is important to mention that the only difference between these two integrators is the order of integrating the last two pieces. In particular, an additional force calculation is needed in the PNHL-N scheme just before updating the last B piece at the end of each integration step. In experiments, the high per-timestep cost of PNHL-N was found to be offset by a great increase in accuracy and usable steplength. Detailed numerical comparisons will be presented in Section 3.7.

3.6 Error Analysis

To our knowledge, little attention has been paid to the mathematical analysis of the DPD system, or more generally stochastic momentum-conserving thermostats. Since the spectral properties of the Fokker–Planck operators in the case of DPD is not available, a rigorous study of the order of convergence of numerical methods in this context has been lacking. Because of the inclusion of stochastic terms, it is not reliable to depend directly on intuition regarding the error behavior of deterministic schemes [66, 105]. Here we perform a few first steps toward the analysis of stochastic DPD integrators by extending a framework for investigating the perturbation of long-time average computed using numerical methods in Langevin dynamics proposed recently by Leimkuhler and Matthews [99, 100].

3.6.1 Expansion of the Invariant Measure

The analysis of the accuracy of ergodic averages (averages with respect to the invariant measure) in stochastic numerical methods can be performed using the framework of long-time Talay–Tubaro expansion, as developed in [1, 2, 38, 99, 100, 102, 163].

For a splitting method described by $\mathcal{L} = \mathcal{L}_\alpha + \mathcal{L}_\beta + \cdots + \mathcal{L}_\zeta$, we define the effective operator $\hat{\mathcal{L}}^\dagger$ associated with the perturbed system obtained using the numerical method with stepsize h by the relation

$$\exp\left(h\hat{\mathcal{L}}^\dagger\right) = \exp\left(h\mathcal{L}_\alpha^\dagger\right) \exp\left(h\mathcal{L}_\beta^\dagger\right) \ldots \exp\left(h\mathcal{L}_\zeta^\dagger\right) . \tag{3.63}$$

This operator can be computed using the Baker–Campbell–Hausdorff (BCH) expansion and can thus be viewed as a perturbation of the exact Fokker–Planck operator $\mathcal{L}^\dagger$:

$$\hat{\mathcal{L}}^\dagger = \mathcal{L}^\dagger + h\mathcal{L}_1^\dagger + h^2\mathcal{L}_2^\dagger + O(h^3) \tag{3.64}$$

for some perturbation operators $\mathcal{L}_i^\dagger$.

We also define the invariant distribution $\hat{\rho}$ associated with the numerical method as an approximation of the target invariant distribution ρ_β:

$$\hat{\rho} = \rho_\beta \left[1 + hf_1 + h^2 f_2 + h^3 f_3 + O(h^4)\right] \tag{3.65}$$

for some correction functions f_i satisfying $\langle f_i \rangle = 0$.

Substituting $\hat{\mathcal{L}}^\dagger$ and $\hat{\rho}$ into the stationary Fokker–Planck equation

$$\hat{\mathcal{L}}^\dagger \hat{\rho} = 0$$

yields

$$\left(\mathcal{L}^\dagger + h\mathcal{L}_1^\dagger + h^2\mathcal{L}_2^\dagger + O(h^3)\right)\left(\rho_\beta \left[1 + hf_1 + h^2 f_2 + h^3 f_3 + O(h^4)\right]\right) = 0\,.$$

Since the exact Fokker–Planck operator preserves the invariant canonical distribution, i.e., $\mathcal{L}^\dagger \rho_\beta = 0$, we obtain

$$\mathcal{L}^\dagger(\rho_\beta f_1) = -\mathcal{L}_1^\dagger \rho_\beta \tag{3.66}$$

by equating first order terms in h.

For any particular integration scheme it is possible to find the perturbation operator $\mathcal{L}_1^\dagger$ by using the BCH expansion: for (noncommutative) linear operators X and Y, we have

$$\exp(hX)\exp(hY) = \exp(hZ_1)\,,$$

where

$$Z_1 = X + Y + \frac{h}{2}[X,Y] + \frac{h^2}{12}([X,[X,Y]] - [Y,[X,Y]]) + O(h^3)\,, \tag{3.67}$$

and subsequently

$$\exp\left(\frac{h}{2}X\right)\exp(hY)\exp\left(\frac{h}{2}X\right) = \exp(hZ_2)\,,$$

where

$$Z_2 = X + Y + \frac{h^2}{12}\left([Y,[Y,X]] - \frac{1}{2}[X,[X,Y]]\right) + O(h^4)\,. \tag{3.68}$$

The notation $[X,Y] = XY - YX$ denotes the commutator of operators X and Y.

Then we can calculate its action on ρ_β. The last step, namely obtaining the leading correction function f_1, requires the solution of the above PDE (3.66) (see examples in Langevin dynamics [99]). In general, solving for f_1 in closed form is difficult, and it does not get simpler for more complicated formulations than Langevin dynamics and more complicated splittings.

These equations demonstrate that for nonsymmetric splitting methods, there typically exists a nonzero term $\mathcal{L}_1^\dagger \propto [X, Y] \neq 0$, while the condition $\mathcal{L}_1^\dagger = 0$, implying $f_1 = 0$, is automatically satisfied for symmetric splitting methods; thus, for observables ϕ, assuming the asymptotic expansion holds, the computed average would be of order two:

$$\langle\phi\rangle_h = \langle\phi\rangle + h\langle\phi f_1\rangle + h^2\langle\phi f_2\rangle + \cdots = \langle\phi\rangle + O(h^2)\,,$$

where $\langle\cdot\rangle$ denotes the average with respect to the target invariant distribution.

3.6.2 Weak and Long-Time Errors

In simulations, the error in averages related to the evolving distribution is generally of interest, i.e., the *weak error* (finite-time error in the weak sense) for nonequilibrium and dynamical properties, and, *long-time error* (error in the invariant distribution obtained as $t \to \infty$) for thermodynamics. We next give definitions of these two errors following [103].

For the weak error, consider a finite time interval $[0, \tau]$ with $\tau = nh$. The probability measure associated with a certain system is described by a probability density $\rho(\mathbf{z}, t)$ which evolves according to the Fokker–Planck equation

$$\frac{\partial\rho}{\partial t} = \mathcal{L}^\dagger\rho\,,$$

where $\mathcal{L}^\dagger$ is the adjoint of the generator of the system (for instance, the Fokker–Planck operator of the standard DPD system is given in [50]). Assuming ergodicity, the solution $\rho(\mathbf{z}, t)$ evolves from an initial probability distribution $\rho(\mathbf{z}, 0)$ to the steady state (invariant distribution) $\rho(\mathbf{z}, \infty) = \rho_\beta$. For a smooth and bounded test function ϕ of a suitable class, the average of ϕ at time τ may be defined by

$$\bar{\phi}(\tau) = \int_0^\tau \phi(\mathbf{z})\rho(\mathbf{z}, \tau)\,\mathrm{d}\mathbf{z}\,. \tag{3.69}$$

The discretization scheme can also be viewed as giving rise to an evolving sequence of probability distributions $\rho_1, \rho_2, \cdots$. With stepsize h, the average at time $\tau = nh$ is defined as

$$\hat{\phi}(\tau, h) = \int_{\Omega_\mathbf{z}} \phi(\mathbf{z})\rho_n(\mathbf{z})\,\mathrm{d}\mathbf{z}\,. \tag{3.70}$$

Thus, we could define the *weak error* as the difference between (3.69) and (3.70)

$$|\bar{\phi}(\tau) - \hat{\phi}(\tau, h)| \leq K(\tau)h^p\,, \tag{3.71}$$

where the coefficient K depends on the time interval and p is the order of the weak error. To be more precise, K also depends on the initial distribution $\rho(\mathbf{z}, 0)$ as well as the particular observable ϕ. The asymptotic ($\tau \to \infty$) behavior [163] of K can be used to describe the performance of the numerical method for computing

averages with respect to the invariant distribution. Hence, the *long-time error* in averages can be written as

$$\lim_{\tau\to\infty} |\bar{\phi}(\tau) - \hat{\phi}(\tau, h)| \leq K(\tau)h^p \,. \tag{3.72}$$

3.6.3 Applications in DPD and Related Systems

According to the BCH formula, a nonsymmetric splitting method generally gives only first order convergence for the invariant measure (*long-time error*), whereas second order is anticipated in symmetric splittings in the asymptotic limit of small stepsize [102]. In our numerical studies, we have verified second order convergence for a number of nonsymmetric methods (DPD-S1, LA, and Peters) which are similar to the family of geometric Langevin algorithms (GLA) following [22]. We compute the *long-time error* of various observables, including kinetic and configurational temperatures and average potential energy, and demonstrate the convergence order for each method in Table 3.1. The results shown in Table 3.1 are based on the numerical experiments in Section 3.7. All the symmetric methods show second order as well as those three nonsymmetric ones (DPD-S1, LA, and Peters). Some other nonsymmetric methods, which are not of GLA type, exhibit second order convergence in calculated quantities; this observation remains to be demonstrated rigorously. It is entirely possible that the superconvergence observed in these special cases is related to the form of the observable we have used in our simulation test.

Method	DPD-VV	DPD-S1	DPD-Trotter	LA	Peters	NHLA	PNH	PNHL-S	PNHL-N
Order	$\geq 1(2)$	2	2	2	2	$\geq 1(2)$	$\geq 1(2)$	2	$\geq 1(2)$

Table 3.1: Orders of accuracy of the *long-time error* for kinetic and configurational temperatures, and, average potential energy in various methods are summarized. Theoretical values based on properties of the discretization scheme have been verified using numerical experiments. Where the theoretically expected result differs from the numerical result, we give the numerically observed convergence order in parentheses.

In what follows, we verify that three nonsymmetric splitting methods (DPD-S1, LA, and Peters) are indeed second order convergence to the invariant measure. In the example of DPD-S1, which can be termed OBAB and clearly is not symmetric, each interacting pair preserves the invariant distribution ρ_β (2.34), i.e.,

$$\mathcal{L}^\dagger_{\mathrm{O}_{i,j}} \rho_\beta = 0 \,.$$

Thus one can easily verify that the operator $\mathcal{L}^\dagger_{\mathrm{O}}$ also preserves the invariant distribution

$$\mathcal{L}^\dagger_{\mathrm{O}} \rho_\beta = 0 \,,$$

since all the actions of commutators in the BCH expansion of $\mathcal{L}^\dagger_{\mathrm{O}}$ on the invariant distribution ρ_β would be zero and therefore all the actions of perturbation opera-

tors $\mathcal{L}_i^\dagger$ in (3.64) would be zero. Also, it can be easily shown that the Hamiltonian operator

$$\mathcal{L}_{\mathrm{H}}^\dagger = \mathcal{L}_{\mathrm{A}}^\dagger + \mathcal{L}_{\mathrm{B}}^\dagger$$

preserves the invariant distribution

$$\mathcal{L}_{\mathrm{H}}^\dagger \rho_\beta = 0\,.$$

Thus, by applying the BCH expansion on the operator of the DPD-S1 scheme (3.34), we obtain

$$\hat{\mathcal{L}}_{\mathrm{DPD-S1}}^\dagger = \mathcal{L}_{\mathrm{A}}^\dagger + \mathcal{L}_{\mathrm{B}}^\dagger + \mathcal{L}_{\mathrm{O}}^\dagger + \frac{h}{2}\left(\left[\mathcal{L}_{\mathrm{A}}^\dagger, \mathcal{L}_{\mathrm{O}}^\dagger\right] + \left[\mathcal{L}_{\mathrm{B}}^\dagger, \mathcal{L}_{\mathrm{O}}^\dagger\right]\right) + O(h^2)\,.$$

Hence,

$$\mathcal{L}_1^\dagger \rho_\beta = \frac{1}{2}\left(\left[\mathcal{L}_{\mathrm{A}}^\dagger, \mathcal{L}_{\mathrm{O}}^\dagger\right] + \left[\mathcal{L}_{\mathrm{B}}^\dagger, \mathcal{L}_{\mathrm{O}}^\dagger\right]\right)\rho_\beta = \frac{1}{2}\left[\mathcal{L}_{\mathrm{H}}^\dagger, \mathcal{L}_{\mathrm{O}}^\dagger\right]\rho_\beta = 0\,,$$

which gives the only solution of the PDE (3.66) to the DPD-S1 scheme

$$f_1 = 0\,.$$

Given that higher order perturbations in (3.65) are not generally equal to zero, we have shown that the nonsymmetric DPD-S1 scheme has second order convergence to its invariant distribution. Similarly, we can also demonstrate that both the Lowe–Andersen and Peters thermostats (in the fashion of BABO) maintain second order convergence to the invariant distribution. Since all the three methods involve a second order symplectic Verlet method for the deterministic part and other actions only on the Ornstein–Uhlenbeck (OU) process, we may refer to them as generalized geometric Langevin algorithms of order two (GLA-2) [22] in the context of stochastic momentum-conserving thermostats. It should be emphasized that the OU process in stochastic momentum-conserving thermostats is pairwise and thus different from the standard setting of Langevin dynamics. Nevertheless, it has been demonstrated that second order is still achieved by the combination of a symplectic method for the deterministic part and an exact solve for the OU process. More discussions regarding the long-time accuracy of the GLA-type methods in Langevin dynamics can be found in [2, 22, 102].

3.7 Numerical Experiments

To investigate the performance of all the numerical methods described in Sections 3.3–3.5, we perform systematic numerical experiments in this section.

3.7.1 Simulation Details

Tests have been carried out by using the standard parameter set [65] that is commonly used in algorithms tests [126, 136, 153, 155, 160, 171]. A system of $N = 500$ identical particles ($m_i = m = 1$) was simulated in a cubic box (length $L = 5$) with periodic boundary conditions (particle density $\rho_{\mathrm{d}} = 4$). The cutoff radius is $r_{\mathrm{c}} = 1$ and $k_{\mathrm{B}}T = 1$. The potential repulsion parameter a_{ij} was set to 25, while dissipative strength γ was chosen as 4.5, leading the value of random strength σ to be 3. It is worthy of mention that we did investigate the influence of other values of the friction coefficient, such as $\gamma = 0.5$ and $\gamma = 40.5$, but not all the results will be presented unless necessary. The initial positions of the particles were independent and identically distributed (i.i.d.) with a uniform distribution over the box, while the initial momenta were i.i.d. normal random variables with mean zero and variance $k_{\mathrm{B}}T$. Verlet neighbor lists [172] were used throughout each method.

In particular, the thermal mass μ in the PNH and PNHL thermostats has to be chosen with care. For PNH thermostat, we used $\mu = 200$ to maintain as good stability as possible for the integration scheme, while $\mu = 10$ and $\tilde{\gamma} = 4.5$ were used in the PNHL thermostat. Since the focus is on DPD simulations, the stochastic randomization frequency Γ in LA and NHLA thermostats was set to be 0.44 as in [81,126] so that similar translational diffusion properties of the fluid were obtained.

3.7.2 Measured Physical Quantities

Static

The kinetic temperature T_{k} appears to be the most popular quantity to be tested in DPD literature, which is defined as

$$k_{\mathrm{B}}T_{\mathrm{k}} = \frac{1}{N_{\mathrm{d}}} \sum_i \frac{\mathbf{p}_i \cdot \mathbf{p}_i}{m_i} \,. \tag{3.73}$$

where N_{d} denotes the number of degrees of freedom of the system. But in practice, the kinetic temperature is not that important, instead those configurational quantities play more crucial roles. Recent studies [27, 149] in simulations showed that the system temperature can be measured from static snapshots of its constituents' instantaneous configurations rather than their momenta. We test the configurational temperature T_{c} [149], which can be defined as

$$k_{\mathrm{B}}T_{\mathrm{c}} = \frac{\sum_i \langle \|\nabla_i U\|^2 \rangle}{\sum_i \langle \nabla_i^2 U \rangle} \,, \tag{3.74}$$

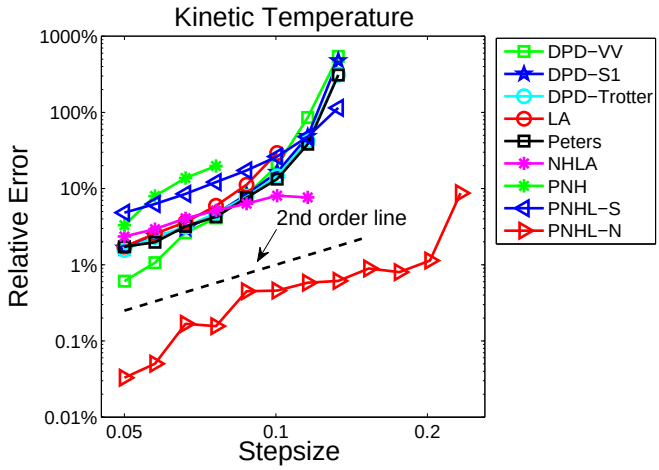


Figure 3.1: Log-log plot of the relative error in computed kinetic temperature against stepsize by using various numerical methods. The system was simulated for 1000 reduced time units but only the last 80% of the data were collected to calculate the static quantity to make sure the system was well equilibrated. The stepsizes tested began at $h = 0.05$ and were increased incrementally by 15% until all methods either were above 100% relative error or became unstable.

where $\nabla_i U$ and $\nabla_i^2 U$ are respectively the gradient and Laplacian of the potential energy U with respect to the position of particle i (see more discussions on the configurational temperature, which, as the kinetic temperature, should in principle be equal to the target temperature, in [27,83,167]). To test the correctness of codes and/or algorithms in simulations, both kinetic and configurational temperatures can be used to calculate the system temperature. However, it turned out that the configurational temperature is more reliable [27,149] since it can rapidly and accurately track changes in system temperature even when the system is not in global thermodynamic equilibrium. It becomes more crucial when it comes to experimental studies of soft condensed matter systems [67,68] most notably due to their applicability to overdamped systems whose instantaneous momenta may not be accessible. It was den Otter and Clarke that first investigated the deviations of the kinetic and configurational temperatures from the system temperature in the DPD system [43]. Since then, little attention has been paid to the configurational temperature in DPD simulations until Allen recently argued that the configurational temperature should be added to the list of diagnostic tests applied to DPD simulations [5]. In addition, we also calculate the average potential energy $\langle U \rangle$ and the radial distribution function (RDF) $g(r)$ [7,58], both of which are very important configurational quantities in simulations.

Dynamical

To have a deeper understanding of how the physical system evolves, it is not enough to just evaluate the static quantities described above. It is essential to measure and compare the relevant dynamical properties. In general, in simula-

tion, one relies on various Green–Kubo formulas [51] to calculate various transport coefficients; in this thesis we restrict ourselves to two special cases.

The velocity autocorrelation function (VAF) is an important measure of dynamical fidelity that numerical methods should be able to reproduce, particularly if they are to be used for nonequilibrium (transient-phase simulation). The VAF, which characterizes the translational diffusion of the system, is defined as

$$\psi(t) = \langle \mathbf{v}_i(t) \cdot \mathbf{v}_i(0) \rangle \,,$$

where $\mathbf{v}_i(0)$ is the initial velocity picked up after the system is well equilibrated. By integrating the VAF (Green–Kubo relation [63, 92]), we can compute the translational diffusion coefficient, which can also be obtained by using Einstein's relation [47] giving the diffusion coefficient as the slope of the mean square displacement as a function of time t:

$$D = \lim_{t\to\infty} \frac{1}{6t} \langle [\mathbf{q}_i(t) - \mathbf{q}_i(0)]^2 \rangle \,.$$

To investigate the rotational relaxation process of the system, we measure the transverse momentum autocorrelation function (TMAF) [69,164], which is related to the shear viscosity, η, in the hydrodynamic limit, i.e., small wavenumber k and large time t, and is defined by

$$C(k,t) = \langle p_x(k,t) p_x(-k,0) \rangle \propto \exp\left(-\frac{k^2\eta}{\rho m} t\right) \,,$$

where ρ is the particle density, m is the particle mass and $p_x(k,t)$ represents the transverse component of the momentum,

$$p_x(k,t) = \sum_{j=1}^{N} p_{j,x}(t) \exp(ikq_{j,z}(t)) \,,$$

where $p_{j,x}$ denotes the x-component of the momentum of particle j and similarly $q_{j,z}$ represents the z-component of its position (see more details in [164]). Note that, i here is the imaginary unit in complex number.

3.7.3 Results

The kinetic temperature control of various methods was tested and shown in Figure 3.1. According to the black dashed second order line in the figure, all the methods seem to have second order convergence to the invariant measure, which verifies the error analysis results on nonsymmetric DPD-S1, LA, and Peters methods in Section 3.6, but is a bit surprising for DPD-VV, NHLA, PNH, and PNHL-N methods that were also based on nonsymmetric splittings.

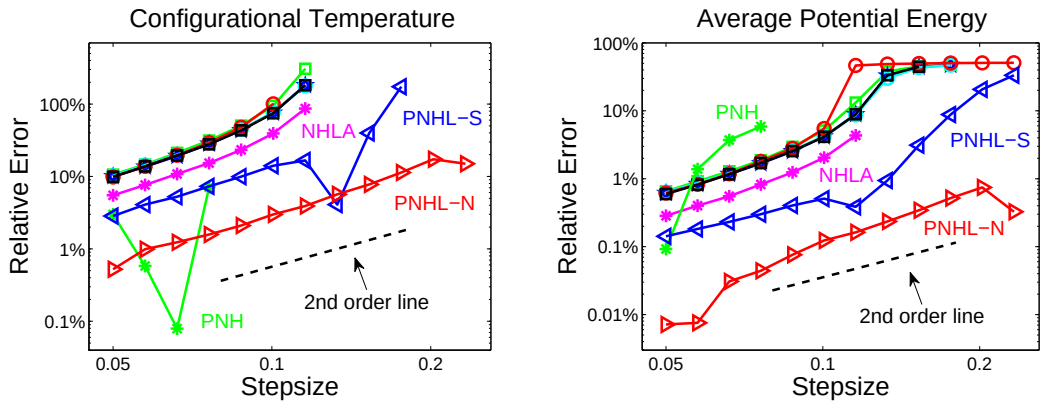


Figure 3.2: Comparisons of the relative error in computed configurational temperature (left) and average potential energy (right) against stepsize by using various numerical methods. The format of the plots is the same as in Figure 3.1. Most of the methods show similar behaviors except NHLA (magenta asterisks), PNH (green asterisks), PNHL-S (blue triangles) and PNHL-N (red triangles).

The performance of standard DPD methods (DPD-VV, DPD-S1, and DPD-Trotter) and the Peters thermostat that converges to the standard DPD system in the limit of $h \to 0$, are almost indistinguishable with the tendency to quickly blow up after the stepsize reaches $h = 0.1$. Both LA and NHLA thermostats show similar behaviors and maximal stepsizes that can be used are limited around $h = 0.1$. The latter displays a better accuracy when stepsize is large due to the additional Nosé–Hoover-like thermostat (3.54). The PNH thermostat illustrates the worst kinetic temperature accuracy and we can also see the clear low stability threshold for the PNH thermostat due to the lack of ergodicity. Most surprisingly and interestingly, the PNHL-S and PNHL-N methods, based on different splitting orders on the same pairwise Nosé–Hoover–Langevin (PNHL) thermostat (2.45), show dramatically different kinetic temperature control: the symmetric PNHL-S method is worse than most of the methods, whereas the nonsymmetric PNHL-N method outstandingly maintains almost two orders of magnitude improvement on the accuracy of kinetic temperature than all the other methods.

Configurational quantities (configurational temperature and average potential energy) were compared in Figure 3.2. Again, all the methods seem to show second order convergence to the invariant measure except the PNH thermostat, which displays a stability threshold of $h = 0.05$. Most of the methods, including standard DPD methods, LA and Peters thermostats, are indistinguishable and cross the 100% barrier in configurational temperature and 10% barrier in average potential energy respectively around $h = 0.1$. As in the case of kinetic temperature, the performance of the NHLA thermostat on those configurational quantities is slightly better than the majority due to the additional thermostat. Unlike the kinetic temperature case, the PNHL-S method does have very good accuracy in configurational quantities with almost one order of magnitude improvement than the majority. Incredibly, the PNHL-N method manages more than one order of magnitude accuracy enhancement in configurational temperature and almost two orders of magnitude in average potential energy.

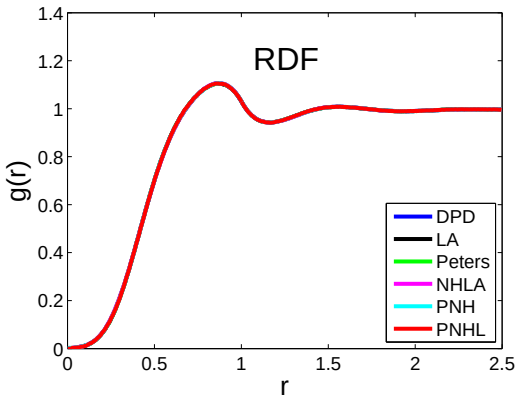


Figure 3.3: Comparisons of radial distribution function (RDF) $g(r)$ of various numerical methods by using very small stepsize $h = 0.01$. All the methods exhibit exactly the same RDF.

Figure 3.3 compares the radial distribution function (RDF) that characterizes the structure of the fluids. The DPD-S1 and PNHL-N methods were used for the standard DPD and PNHL systems respectively. Given that very small stepsizes were used, all the methods display exactly the same RDF, which indicates that different systems maintain the same structure of the fluids without the impacts of discretization errors. Expectedly, discretization errors start to destroy the structure of the fluids with increasing stepsizes as shown in Figure 3.4. Standard DPD methods and the Peters thermostat exhibit similar behaviors with the RDFs starting to show artifacts around $h = 0.09$ and being heavily destroyed around $h = 0.13$. The LA and PNH thermostats again show lower stability in maintaining the correct structure, blowing up around $h = 0.11$ and 0.07, respectively. The performance of PNHL-S and NHLA methods are slightly better than the majority of the schemes, while the PNHL-N method only starts to show pronounced artifacts above $h = 0.25$, which is remarkably more than two times larger than the stepsize usable in standard DPD methods.

The velocity autocorrelation function (VAF) of various numerical methods were calculated in Figure 3.5 to compare with standard DPD methods with three different values of friction coefficient. The DPD-S1 method was used to calculate the VAF of standard DPD methods since there is no difference with other two if very small stepsizes were used to reduce the effects of discretization errors. Similarly, the uncorrupted dynamics of PNHL-S and PNHL-N methods should be exactly the same, and the latter was used. The VAFs of the PNH and PNHL systems are indistinguishable and consistent with standard DPD methods in the regime of small friction coefficient $\gamma = 0.5$. This is not surprising since the average of the dynamical variable ξ, which is controlled by an additional thermostat and can be thought of as the "dynamical friction", tends to zero. As we expected, the Peters thermostat shares the same VAF as standard DPD methods in the regime of the commonly used friction coefficient $\gamma = 4.5$ if the same value of γ was chosen. The stochastic randomization frequency $\Gamma = 0.44$ was used in LA and NHLA thermostats to maintain similar translational diffusion properties as

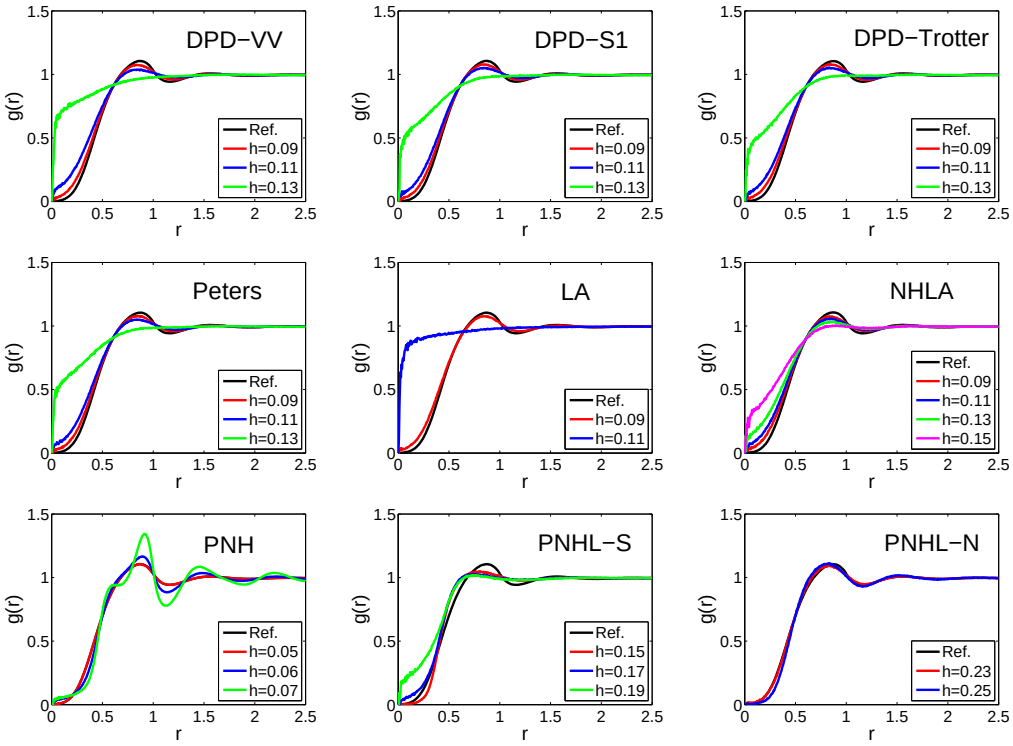


Figure 3.4: Stepsize effects on the radial distribution function (RDF) in various numerical methods were compared. The black solid lines are the reference solutions obtained using very small stepsizes $h = 0.01$, while the colored lines correspond to different stepsizes.

in the standard DPD system with $\gamma = 4.5$. The VAF of the standard DPD system with large friction coefficient $\gamma = 40.5$ is also shown in the figure, which indicates that the larger the friction is the faster the VAF goes down (the system loses memory faster).

The effects of the discretization error on each method were also investigated in Figure 3.6. The results are largely consistent with those observed for the configurational temperature. The only surprise is that the PNHL-S method allows a maximum stepsize which is similar to those of the thermostats considered (and much lower than the useful stepsize for PNHL-N). Moreover, we observe that the PNH thermostat begins to display unphysical artifacts, at a stepsize of just $h = 0.05$. Likewise, the LA thermostat has a lower stability threshold. Among the various schemes, PNHL-N is again by far the most stable scheme, exhibiting only a mild deviation from the reference VAF at $h = 0.17$.

One of the most important features of DPD simulations is the correct handling of rotational relaxation which will be important for resolving correct vortical motion and long-ranged interactions. We therefore investigate the computation of transverse momentum autocorrelation function (TMAF, see Section 3.7.2) for each scheme. The results here are in many ways similar to those obtained for the velocity autocorrelation function (Figure 3.5), thus only the results of DPD

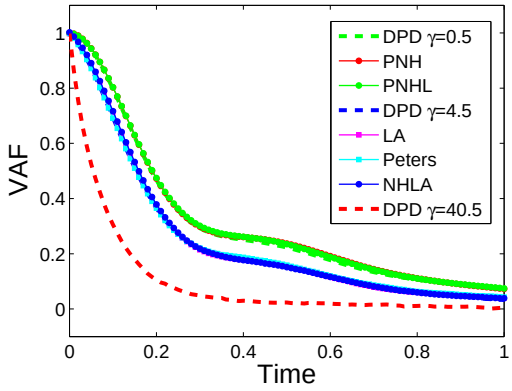


Figure 3.5: Comparisons of velocity autocorrelation function (VAF) of various numerical methods by using very small stepsize $h = 0.01$. Standard DPD methods with three different values of friction coefficient were calculated by using the DPD-S1 method to compare with other methods. 100 different runs were averaged to reduce the sampling errors after the system was well equilibrated.

and PNHL methods are shown in Figure 3.7: the PNHL method is expectedly consistent with small friction ($\gamma = 0.5$) in standard DPD simulation, and only a minor difference is observed between the PNHL method and standard DPD with a moderate value of friction ($\gamma = 4.5$). One may notice that the TMAF shown in Figure 3.7 is not as smooth as the VAF in Figure 3.5 even by averaging 1000 times more different runs. We emphasize here that collective (N-particle) correlations, such as TMAF and stress autocorrelation function [31], fluctuate rapidly and thus are always determined with poorer statistics than single-particle ones, such as VAF.

We emphasize the following facts regarding our numerical tests:

- A detailed statistical analysis of the results presented has not been incorporated in this thesis, due to the extensive computational requirements of doing so. An early study [155] has already suggested that highly reliable estimation of the kinetic temperature can be typically obtained by various methods in the stepsize regime of our interests, i.e., $h \geq 0.05$. In terms of convergence rate of thermodynamic properties to distribution, all the methods perform similarly in practice (see Figures 3.1–3.2).

- The influence of the friction coefficients (three different values, $\gamma = 0.5$, 4.5, and 40.5), both on the maximal timestep and the dynamics of the system, has been investigated in this section. Different values of the friction coefficients gave little difference in the control of those static quantities we calculated, thus only the results of $\gamma = 4.5$ have been presented. However, the dynamical properties of the system did appear to depend on the strength of the friction as shown in the velocity and transverse momentum autocorrelation functions.

- Based on the results for the VAF and TMAF, we have observed very good

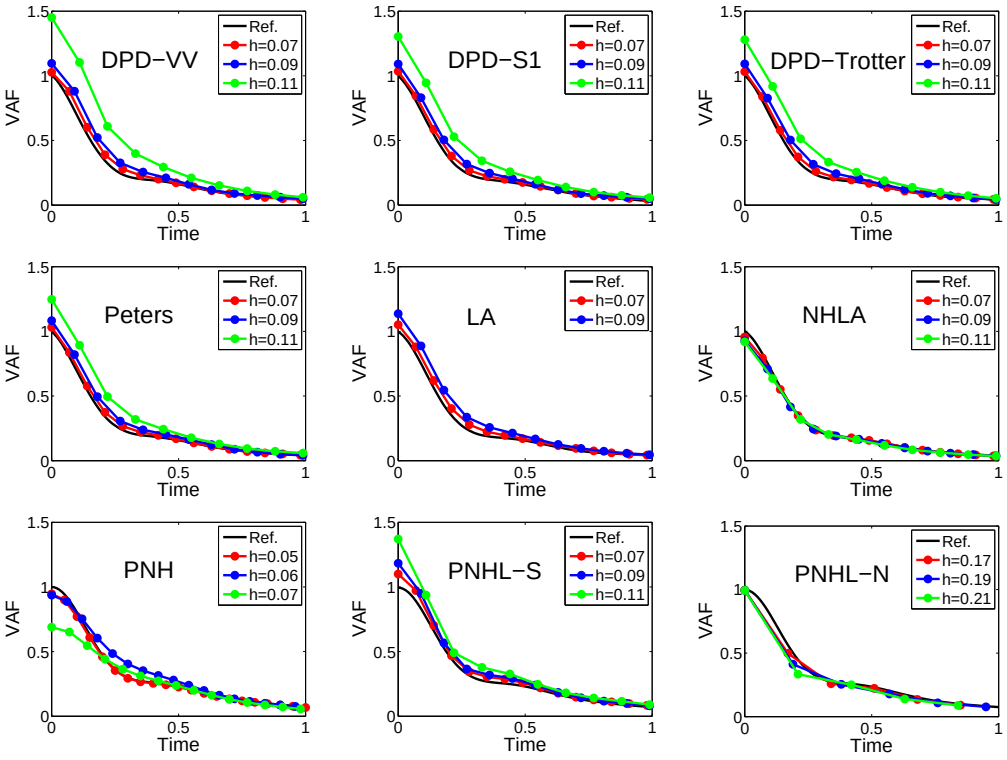


Figure 3.6: Stepsizes effects on the velocity autocorrelation function (VAF) in various numerical methods were compared. The black solid lines are the reference solutions obtained using very small stepsizes, while those colored lines correspond to different stepsizes. The error in the VAFs at time zero reflects the bias, visible at large stepsize, in the mean kinetic energy (which is the normalization factor). The VAFs could be rescaled so that they start from one, but it would mask the presence of this large disturbance from the target temperature.

agreement of the dynamical properties between PNHL and DPD with relatively small friction coefficient, particularly $\gamma = 0.5$. Although the PNHL method may not be able to fully recover the hydrodynamics as DPD, we have already seen that PNHL offers substantially improved stability in simulations. Moreover, the PNHL method, as a general momentum-conserving thermostat, has a valid motivation in a broader context than just comparison with DPD; in particular it will be useful in NEMD and in other cases where fluid dynamics per se is not at issue.

3.7.4 Computational Efficiency

The computational efficiency of the various methods was tested with simulation details in Section 3.7.1. All the tests were run on an HP Z600 Workstation with 15.7 GB RAM. As shown in Table 3.2, we calculated the CPU time (milliseconds) taken with the use of Verlet neighbor lists for the integration of a single time step of $h = 0.05$ (averaged over 10,000 consecutive time steps). Note that only the

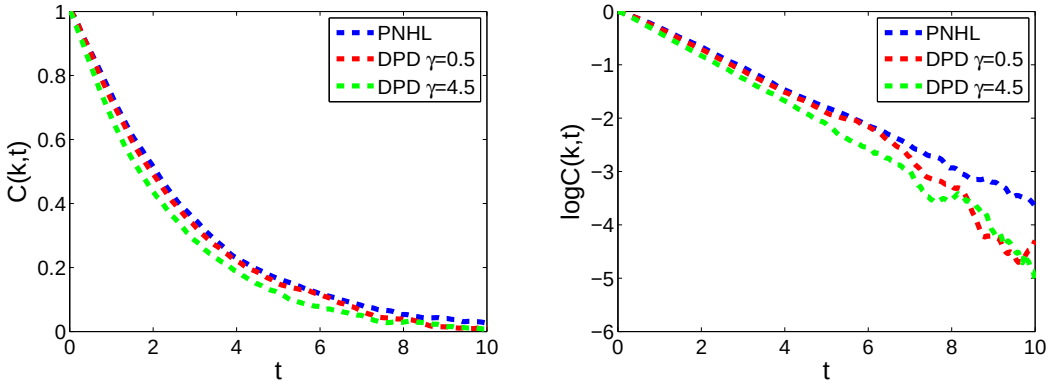


Figure 3.7: (Color) Comparisons of the normalized transverse momentum autocorrelation function (left) and its logarithm (right) between standard DPD and PNHL methods with stepsize $h = 0.05$. DPD-S1 and PNHL-N were used to solve the DPD and PNHL systems respectively. The ratio of the curves on the right panel is proportional to the corresponding shear viscosity. The wavenumber k was chosen as $2\pi/L$ and 100,000 different runs were averaged to reduce the sampling errors after the system was well equilibrated.

time for the main integration step (without calculating any physical quantity) was counted.

Method	**Critical Stepsize**	**Maximal Stepsize**	**CPU Time**	**Scaled Efficiency**
DPD-VV	0.05	0.10	17.710	100.0%
DPD-S1	0.05	0.11	17.783	99.6%
DPD-Trotter	0.05	0.11	18.482	95.8%
LA	0.05	0.10	15.364	115.3%
Peters	0.05	0.11	18.286	96.9%
NHLA	0.07	0.13	16.281	152.3%
PNH	0.05	0.08	13.990	126.6%
PNHL-S	0.08	0.17	19.408	146.0%
PNHL-N	0.17	0.23	32.183	187.1%

Table 3.2: Comparisons of the computational efficiency of the various numerical methods. "Critical stepsize" is the stepsize beyond which the numerical method starts to show pronounced artifacts, while "maximal stepsize" is the stepsize stability threshold. The "numerical efficiency" of each method was scaled to that of the benchmark DPD-VV method.

In order to quantitatively compare the overall performance of each method, we define a quantity, the "numerical efficiency", that measures the amount of simulation time accessible per unit of computational work, i.e.,

$$\text{Numerical Efficiency} = \frac{\text{Critical Stepsize}}{\text{CPU Time Per Step}}, \tag{3.75}$$

where the "critical stepsize" is defined as the stepsize beyond which pronounced

artifacts become apparent in some observable thus rendering the simulation unusable. For practical purposes, we use a threshold of 10% relative error in configurational temperature in determining the critical stepsize. Our reasoning in choosing the configurational temperature as the quantity for determination of the critical stepsize is that (a) it is a sensitive observable and difficult to control in simulation, (b) accuracy of configurational temperature seems to us to be intuitively to be a core requirement for canonical simulation methods, (c) good control of the configurational temperature appears to imply good results in other stationary computations. The choice of 10% as the bar for accuracy is clearly arbitrary. If a smaller error threshold were used, the results may change slightly, but the ordering of the methods in terms of efficiency would remain essentially the same.

We also show the "maximal stepsize" in Table 3.2, which is the stepsize at which the numerical method is either above 100% relative error in configurational temperature or unstable. Only the PNHL-N method needs to calculate the force twice in each integration step, which is the reason why an almost doubled time was needed than other methods. Finally, the computed "numerical efficiency" of each method was scaled to that of the DPD-VV method since it is still the most popular method in DPD simulations and we use that method as the benchmark in the comparisons.

As can be seen from the table, those standard DPD methods and the Peters thermostat have comparable "numerical efficiency". The LA and PNH thermostats are slightly better than the commonly used DPD-VV method with around 15% and 27% improvements, respectively. Both NHLA and PNHL-S methods maintain an about 50% enhancement. The improvement of the "numerical efficiency" of the PNHL-N is remarkably 87% in comparison with the DPD-VV method. Similarly, we may also compute the enhancements of the PNHL-N method based on kinetic temperature and average potential energy, using the same critical stepsize $h = 0.05$ as in configurational temperature: 43% for the former and 120% for the latter.

3.8 Second Order Convergence of PNHL-N for Certain Observables

We have demonstrated in Section 3.6.1 that a symmetric splitting method would give at least second order convergence to the invariant measure. We have also showed that nonsymmetric splitting methods of GLA type would have second order convergence in Section 3.6.3. However, although second order convergence was also observed for the nonsymmetric splitting method of PNHL (i.e., the PNHL-N method) for a number of observables in Section 3.7, its theoretical verification has been lacking. In this section, by adopting the procedures based on the expansion of the invariant measure outlined in Section 3.6.1, we verify the

second order convergence of the PNHL-N method for certain observables.

Consider the PNHL-N method (3.62), which can be written as

$$\exp\left(h\hat{\mathcal{L}}^{\dagger}_{\mathrm{PNHL-N}}\right) = \exp\left(\frac{h}{2}\mathcal{L}^{\dagger}_{\mathrm{X}}\right)\exp\left(h\mathcal{L}^{\dagger}_{\mathrm{Y}}\right)\exp\left(\frac{h}{2}\mathcal{L}^{\dagger}_{\mathrm{X}}\right)\,,$$

where

$$\exp\left(\frac{h}{2}\mathcal{L}^{\dagger}_{\mathrm{X}}\right) = \exp\left(\frac{h}{2}\mathcal{L}^{\dagger}_{\mathrm{B}}\right)\exp\left(\frac{h}{2}\mathcal{L}^{\dagger}_{\mathrm{A}}\right)\,,$$

and

$$\exp\left(h\mathcal{L}^{\dagger}_{\mathrm{Y}}\right) = \exp\left(\frac{h}{2}\mathcal{L}^{\dagger}_{\mathrm{C}}\right)\exp\left(\frac{h}{2}\mathcal{L}^{\dagger}_{\mathrm{D}}\right)\exp\left(h\mathcal{L}^{\dagger}_{\mathrm{O}}\right)\exp\left(\frac{h}{2}\mathcal{L}^{\dagger}_{\mathrm{D}}\right)\exp\left(\frac{h}{2}\mathcal{L}^{\dagger}_{\mathrm{C}}\right)\,.$$

By using the BCH expansion, we obtain

$$\begin{aligned}
\mathcal{L}^{\dagger}_{\mathrm{X}} &= \mathcal{L}^{\dagger}_{\mathrm{A}} + \mathcal{L}^{\dagger}_{\mathrm{B}} - \frac{h}{4}\left[\mathcal{L}^{\dagger}_{\mathrm{A}}, \mathcal{L}^{\dagger}_{\mathrm{B}}\right] + O(h^2)\,,\\
\mathcal{L}^{\dagger}_{\mathrm{Y}} &= \mathcal{L}^{\dagger}_{\mathrm{C}} + \mathcal{L}^{\dagger}_{\mathrm{D}} + O(h^2)\,,
\end{aligned}$$

and subsequently

$$\hat{\mathcal{L}}^{\dagger}_{\mathrm{PNHL-N}} = \mathcal{L}^{\dagger}_{\mathrm{A}} + \mathcal{L}^{\dagger}_{\mathrm{B}} + \mathcal{L}^{\dagger}_{\mathrm{C}} + \mathcal{L}^{\dagger}_{\mathrm{D}} + \mathcal{L}^{\dagger}_{\mathrm{O}} - \frac{h}{4}\left[\mathcal{L}^{\dagger}_{\mathrm{A}}, \mathcal{L}^{\dagger}_{\mathrm{B}}\right] + O(h^2)\,.$$

Thus the leading perturbation operator of the PNHL-N method is

$$\mathcal{L}^{\dagger}_{1,\mathrm{PNHL-N}} = -\frac{1}{4}\left[\mathcal{L}^{\dagger}_{\mathrm{A}}, \mathcal{L}^{\dagger}_{\mathrm{B}}\right]\,, \tag{3.76}$$

whose action on the invariant distribution of the PNHL system, i.e.,

$$\hat{\rho}_{\beta}(\mathbf{q},\mathbf{p},\xi) = \frac{1}{Z}\exp\left(-\beta H(\mathbf{q},\mathbf{p})\right)\exp\left(-\frac{\beta\mu\xi^2}{2}\right)\,, \tag{3.77}$$

reads (assuming $\mathbf{M} = \mathbf{I}$ for simplicity)

$$\mathcal{L}^{\dagger}_{1,\mathrm{PNHL-N}}\hat{\rho}_{\beta}(\mathbf{q},\mathbf{p},\xi) = -\frac{\beta}{4}\left(\mathbf{p}^T\Delta U(\mathbf{q})\mathbf{p} - [\nabla U(\mathbf{q})]^T\nabla U(\mathbf{q})\right)\hat{\rho}_{\beta}\,. \tag{3.78}$$

Although in this case the right-hand side of the PDE (3.66) is relatively simple, it is still very challenging to solve the PDE in order to obtain the corresponding leading correction function $f_{1,\mathrm{PNHL-N}}$. However, the additional variable ξ is normally distributed with mean zero and variance $\beta^{-1}\mu^{-1}$. Thus, the variance of ξ will be small if the thermal mass μ is large. Therefore, we consider projecting

the Fokker–Planck equation, which reads

$$\mathcal{L}^\dagger_{\mathrm{PNHL}} = -\mathbf{p}\cdot\nabla_{\mathbf{q}} + \nabla U(\mathbf{q})\cdot\nabla_{\mathbf{p}} + \xi\nabla_{\mathbf{p}}\cdot(\mathbf{\Gamma}(\mathbf{q})\mathbf{p}\cdot) - G(\mathbf{q},\mathbf{p})\frac{\partial}{\partial\xi} + \tilde{\gamma}\frac{\partial}{\partial\xi}(\xi\cdot) + \frac{\tilde{\sigma}^2}{2}\frac{\partial^2}{\partial\xi^2}\,,$$

and its solution by integrating with respect to the Gaussian distribution of ξ in the ergodic limit. That is, we apply the projection operator [61]

$$\mathcal{P}\nu(\mathbf{q},\mathbf{p},\xi) := \frac{\int_{\Omega_\xi}\hat{\rho}_\beta(\mathbf{q},\mathbf{p},\xi)\nu(\mathbf{q},\mathbf{p},\xi)\,\mathrm{d}\xi}{\int_{\Omega_\xi}\hat{\rho}_\beta(\mathbf{q},\mathbf{p},\xi)\,\mathrm{d}\xi}\,, \tag{3.79}$$

where ν is an arbitrary function, to the PDE (3.66). Effectively, this leads to the reduced equation

$$\check{\mathcal{L}}^\dagger(\rho_\beta\hat{f}_1) = -\rho_\beta\mathcal{P}\frac{\mathcal{L}^\dagger_1\hat{\rho}_\beta}{\hat{\rho}_\beta}\,, \tag{3.80}$$

where the operator $\check{\mathcal{L}}^\dagger$, acting on functions of $\mathbf{q}$ and $\mathbf{p}$ only, is just the operator $\mathcal{L}^\dagger$ reduced by the action of the projection. In fact, $\mathcal{L}^\dagger_{\mathrm{PNHL}}$ is now simply reduced to

$$\check{\mathcal{L}}^\dagger_{\mathrm{PNHL}} = -\mathbf{p}\cdot\nabla_{\mathbf{q}} + \nabla U(\mathbf{q})\cdot\nabla_{\mathbf{p}}\,,$$

while the right-hand side is unchanged due to the fact that ξ is not present. Finally, we can solve the reduced PDE to obtain the leading correction function:

$$\hat{f}_{1,\mathrm{PNHL-N}} = \frac{\beta}{4}\mathbf{p}^T\nabla U(\mathbf{q})\,, \tag{3.81}$$

which suggests that the PNHL-N method could have second order convergence to its invariant measure for certain observables, (i.e., in the form of $\phi(\mathbf{q},\mathbf{p}) = \mathbf{p}^{2k}\vartheta(\mathbf{q})$, where k is an integer and $\vartheta(\mathbf{q})$ can be constant), including kinetic temperature and observables that only depend on the configurations. In other words, for those observables, assuming the asymptotic expansion holds, the computed average (in the large thermal mass limit) reads

$$\langle\phi\rangle_h = \langle\phi\rangle + h\langle\phi\hat{f}_1\rangle + h^2\langle\phi\hat{f}_2\rangle + \cdots = \langle\phi\rangle + O(h^2)\,,$$

which is of order two. This is fully consistent with what we have observed numerically for a number of observables in Section 3.7.

Chapter 4

Pairwise Adaptive Langevin (PAdL) Thermostat with Applications in Nonequilibrium Modelling

We have introduced, in Section 2.6, the pairwise Nosé–Hoover–Langevin (PNHL) thermostat (2.45), which acts as an alternative approach to DPD. Two different splitting methods have been proposed in Section 3.5.2 for the PNHL system, with the first being of a symmetric manner, labeled as PNHL-S (3.61), and the other nonsymmetric, PNHL-N (3.62). Both PNHL methods have been systematically compared to various popular schemes presented for a number of physical quantities in Section 3.7, and it turns out that both methods achieve significant improvements in terms of accuracy, robustness, and numerical efficiency over alternatives. However, it appears that the dynamical properties of the PNHL formulation corresponds to that of the standard DPD system in the low friction regime. In this chapter, we propose a new momentum-conserving thermostat in order to have full control of the dynamics in practice. The results presented in this chapter have not appeared anywhere else.

4.1 Formulation of PAdL

We propose here a momentum-conserving pairwise adaptive Langevin (PAdL) thermostat (as a potential replacement for DPD), whose equations of motion (for particle i) are given by

$$\begin{aligned} \mathrm{d}\mathbf{q}_i &= m_i^{-1}\mathbf{p}_i \mathrm{d}t\,, \\ \mathrm{d}\mathbf{p}_i &= \mathbf{F}_i^{\mathrm{C}}(\mathbf{q})\mathrm{d}t - \xi \mathbf{V}_i(\mathbf{q},\mathbf{p})\mathrm{d}t + \sigma \mathbf{R}_i(\mathbf{q},t)\,, \\ \mathrm{d}\xi &= G(\mathbf{q},\mathbf{p})\mathrm{d}t\,, \end{aligned} \tag{4.1}$$

where σ is a constant coefficient. In fact, the PAdL system can be viewed as a "perturbed pairwise Nosé–Hoover (PNH)" thermostat, since the only difference comparing to the PNH thermostat is the additional stochastic term. An additional Langevin thermostat could also be added on to the auxiliary variable ξ as in PNHL in order to improve ergodicity, but it is not that necessary here due to the presence of the additional stochastic term.

It can be easily demonstrated that the PAdL system also preserves the canonical ensemble with a modified invariant distribution

$$\tilde{\rho}_\beta(\mathbf{q},\mathbf{p},\xi) = \frac{1}{Z}\exp\left(-\beta H(\mathbf{q},\mathbf{p})\right)\exp\left(-\frac{\beta\mu}{2}(\xi-\hat{\gamma})^2\right), \tag{4.2}$$

where $\hat{\gamma}$ can be thought of as the "effective friction" and the following fluctuation-dissipation theorem is satisfied as in DPD

$$\sigma^2 = 2\hat{\gamma}k_{\mathrm{B}}T\,. \tag{4.3}$$

The invariant distribution (4.2) implies that the auxiliary variable ξ is Gaussian distributed with mean $\hat{\gamma}$ and variance $\beta^{-1}\mu^{-1}$. The auxiliary variable will fluctuating around its mean value during simulation. Therefore, we can think of the PAdL thermostat as the standard DPD system with an "effective friction". In the large thermal mass limit (i.e., $\mu\to\infty$), the PAdL thermostat effectively reduces to the standard DPD formulation (2.37).

Once again, the newly proposed PAdL thermostat (4.1) can be written in a more compact form

$$\begin{aligned}
\mathrm{d}\mathbf{q} &= \mathbf{M}^{-1}\mathbf{p}\mathrm{d}t\,,\\
\mathrm{d}\mathbf{p} &= -\nabla U(\mathbf{q})\mathrm{d}t - \xi\boldsymbol{\Gamma}(\mathbf{q})\mathbf{M}^{-1}\mathbf{p}\mathrm{d}t + \sigma\boldsymbol{\Sigma}(\mathbf{q})\mathrm{d}\mathbf{W}\,,\\
\mathrm{d}\xi &= G(\mathbf{q},\mathbf{p})\mathrm{d}t\,.
\end{aligned} \tag{4.4}$$

A symmetric splitting scheme for the PAdL thermostat is given in Appendix A: as in the DPD-Trotter scheme (Section 3.3.3), the vector field of the "Ornstein–Uhlenbeck" part of the PAdL system is split into each interacting pair in such a way that we can solve the interaction of each pair exactly.

4.2 Numerical Experiments

In this section, we conduct various numerical experiments, in both equilibrium and nonequilibrium regimes, to compare the newly proposed PAdL method with a number of alternative momentum-conserving schemes described in Chapter 3.

Same simulation details and measured physical quantities were used as in

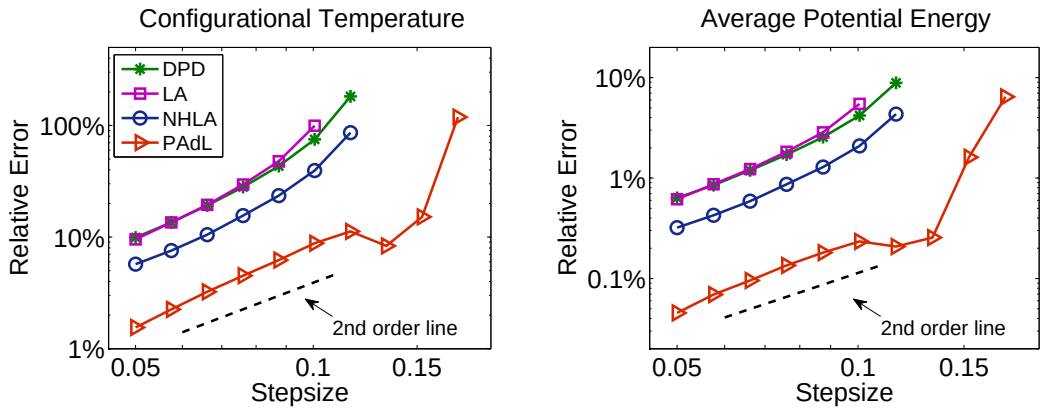


Figure 4.1: Log-log plot of the relative error in computed configurational temperature (left) and average potential energy (right) against stepsize by using various numerical methods with (effective) friction coefficient $\gamma = 4.5$. The system was simulated for 1000 reduced time units but only the last 80% of the data were collected to calculate the static quantity to make sure the system was well equilibrated. Ten different runs were averaged to further reduce the sampling errors. The stepsizes tested began at $h = 0.05$ and were increased incrementally by 15% until all methods either started to show significant relative error (100% in configurational temperature or 10% in average potential energy) or became unstable.

Section 3.7. In particular, the thermal mass in PAdL was chosen to be the same as that of PNHL, i.e., $\mu = 10$. When comparing different formulations, we have to make sure that similar translational diffusion properties of the fluid were obtained. For the PAdL thermostat, we can always tune the value of σ so that the same (effective) friction coefficient as in DPD was obtained, i.e., $\hat{\gamma} = \gamma$. For LA and NHLA thermostats, the stochastic randomization frequency Γ was set to be 0.44 as in [81,126], which corresponds to the case of $\gamma = 4.5$ in DPD. We have also used $\Gamma = 0.1$ and $\Gamma = 4$ for $\gamma = 0.5$ and $\gamma = 40.5$, respectively, in DPD.

We have observed in Section 3.7 that standard DPD methods (DPD-VV, DPD-S1, and DPD-Trotter) and the Peters thermostat give almost indistinguishable performance in all the quantities that we have tested. Therefore, only the DPD-S1 method was used to represent the standard DPD formulation unless otherwise stated. The PNH method was not included for further comparison due to its stability issue, which is well documented in Section 3.7. Since the dynamics of the PNHL thermostat is consistent with that of DPD in the low friction regime, both PNHL methods were only compared to alternatives in the case of $\gamma = 0.5$.

4.2.1 Equilibrium

Configurational quantities, such as configurational temperature and average potential energy, were compared in Figure 4.1 with (effective) friction coefficient $\gamma = 4.5$. According to the dashed order line, we can see that all the methods tested exhibit second order convergence to the invariant measure for both

Method	Critical Stepsize	Maximal Stepsize	CPU Time	Scaled Efficiency
DPD-VV	0.05	0.10	19.878	100.0%
DPD-S1	0.05	0.11	20.018	99.3%
DPD-Trotter	0.05	0.11	20.788	95.6%
Peters	0.05	0.11	20.893	95.1%
LA	0.05	0.10	17.808	111.6%
NHLA	0.07	0.13	18.513	150.3%
PAdL	0.13	0.17	22.219	232.6%

Table 4.1: Comparisons of the computational efficiency of various numerical methods in the moderate (effective) friction regime of $\gamma = 4.5$. "Critical stepsize" is the stepsize beyond which the numerical method starts to show pronounced artifacts, while "maximal stepsize" is the stepsize stability threshold. The "numerical efficiency", defined in (3.75), of each method was scaled to that of the benchmark DPD-VV method.

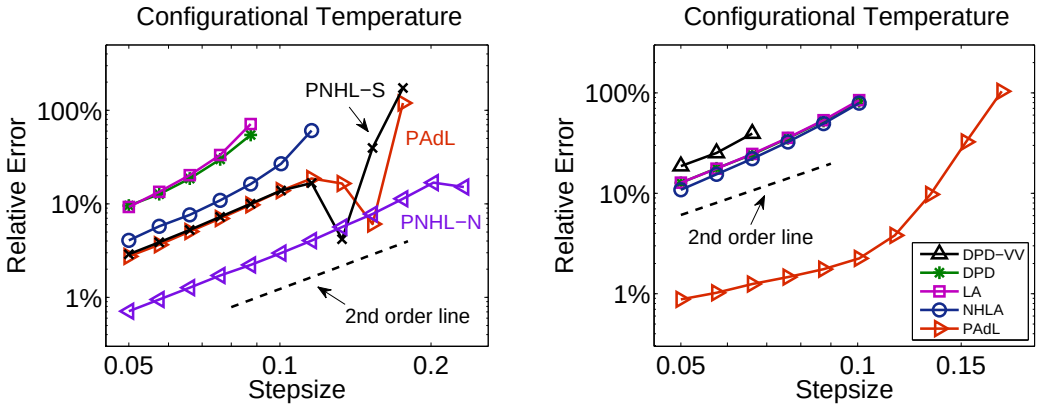


Figure 4.2: Comparisons of the relative error in computed configurational temperature against stepsize by using various numerical methods with (effective) friction coefficient $\gamma = 0.5$ (left) and $\gamma = 40.5$ (right). Note that two pairwise Nosé–Hoover–Langevin (PNHL) methods, which correspond to the low friction regime, were included in the latter case only. The format of the plots is the same as in Figure 4.1.

quantities. More specifically, DPD and the LA thermostat show rather similar behavior, while the NHLA thermostat is slightly better than those two. Quite remarkably, the newly proposed PAdL method achieves one order of magnitude improvement over DPD in terms of numerical accuracy for a fixed stepsize. For certain accuracy (i.e., a fixed relative error), the PAdL method can use doubled stepsize, thus substantially improving the "numerical efficiency" defined in Section 3.7.4. Our observations were confirmed in Table 4.1, which shows that the PAdL method indeed has a more than 130% improvement in numerical efficiency over the DPD method. The results on the configurational temperature and average potential energy are rather similar, therefore in what follows we only present configurational temperature results.

Figure 4.2 (left) compares the configurational temperature control of various methods in both low and high friction regimes. In the low friction regime, where

Method	Critical Stepsize	Maximal Stepsize	CPU Time	Scaled Efficiency
DPD-VV	0.05	0.09	19.878	100.0%
DPD-S1	0.05	0.09	20.018	99.3%
DPD-Trotter	0.05	0.09	20.788	95.6%
Peters	0.05	0.09	20.893	95.1%
LA	0.05	0.09	17.808	111.6%
NHLA	0.07	0.11	18.513	150.3%
PNH	0.05	0.08	16.450	120.8%
PNHL-S	0.08	0.17	21.199	150.0%
PNHL-N	0.17	0.23	37.206	181.7%
PAdL	0.08	0.17	22.219	143.1%

Table 4.2: Comparisons of the computational efficiency of various numerical methods in the low (effective) friction regime of $\gamma = 0.5$. The format of the table is the same as in Table 4.1.

both PNHL methods were included, again all the methods display second order convergence to the invariant measure. The NHLA, PNHL-S, and PAdL methods are rather similar to each other, all of which are superior to both DPD and LA methods. The PNHL-N method achieves more than one order of magnitude improvement in numerical accuracy over the DPD method. Although the PNHL-N method requires two force calculations at each step, it still achieves a more than 80% improvement as shown in Table 4.2. The table also reveals that the PAdL method has an almost 50% improvement over the DPD method.

Method	Critical Stepsize	Maximal Stepsize	CPU Time	Scaled Efficiency
DPD-VV	0.04	0.07	19.878	100.0%
DPD-S1	0.05	0.11	20.018	124.1%
DPD-Trotter	0.05	0.11	20.788	119.5%
Peters	0.05	0.11	20.893	118.9%
LA	0.05	0.11	17.808	139.5%
NHLA	0.05	0.11	18.513	134.2%
PAdL	0.13	0.17	22.219	290.8%

Table 4.3: Comparisons of the computational efficiency of various numerical methods in the high (effective) friction regime of $\gamma = 40.5$. The format of the table is the same as in Table 4.1.

In the high friction regime, the behavior of those methods are rather different from that of other regimes. As shown in Figure 4.2 (right), surprisingly the most popular DPD-VV method is slightly worse than other standard DPD methods. Both LA and NHLA are indistinguishable from the DPD method. Superconvergence property (i.e., fourth order convergence to the invariant measure as demonstrated in Section 5.3.2) was not observed for the PAdL method in this regime. Nevertheless, the PAdL method still obtains a dramatic improvement over all

other schemes. Table 4.3 shows that the PAdL method has a more than 190% improvement in overall numerical efficiency over the benchmark DPD method.

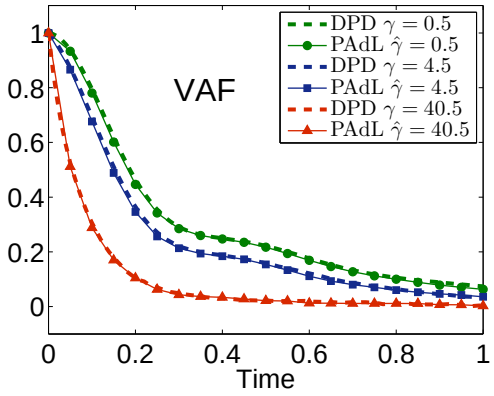


Figure 4.3: Comparisons of velocity autocorrelation function (VAF) between the standard DPD method and the newly established PAdL scheme with three different values of (effective) friction coefficient. The DPD-S1 method was used for DPD with a small stepsize of $h = 0.01$, while $h = 0.05$ was used for PAdL. 100 different runs were averaged to reduce the sampling errors after the system was well equilibrated.

The control of the dynamics of the PAdL method was also tested and plotted in Figure 4.3. Unlike the PNHL formulation, which corresponds to the low friction regime, the PAdL system is able to capture the dynamics of DPD in a wide range of friction coefficients: the relaxations of the velocity autocorrelation function of both formulations are indeed indistinguishable (only visible with the help of the symbol signs).

4.2.2 Nonequilibrium

Before we analyse the numerical results obtained by simulating various methods under Lees–Edwards boundary conditions (i.e., the system is perturbed by a simple shear flow), we discuss two important issues in nonequilibrium molecular dynamics (NEMD) simulations: (1) the practical implementation of the Lees–Edwards boundary conditions in DPD and related momentum-conserving systems, where forces are dependent on relative velocities; (2) the practical measurement of system temperature in NEMD.

A recent paper [30] claimed that, owing to the dependence of inter-particle forces on the relative velocity of the particles, it is problematic to directly apply the LEBC to DPD, especially in high friction regime. As shown in Figure 4.4 (left) where exactly the same setting as in the original paper [30] was used, as the friction increases, the velocity profile starts to (significantly) deviate away from the target linear profile. A simple remedy, which suggests to switch off the interactions of dissipative and random forces (i.e., the DPD thermostat) if one particle is within interacting range of an image of other particle near the

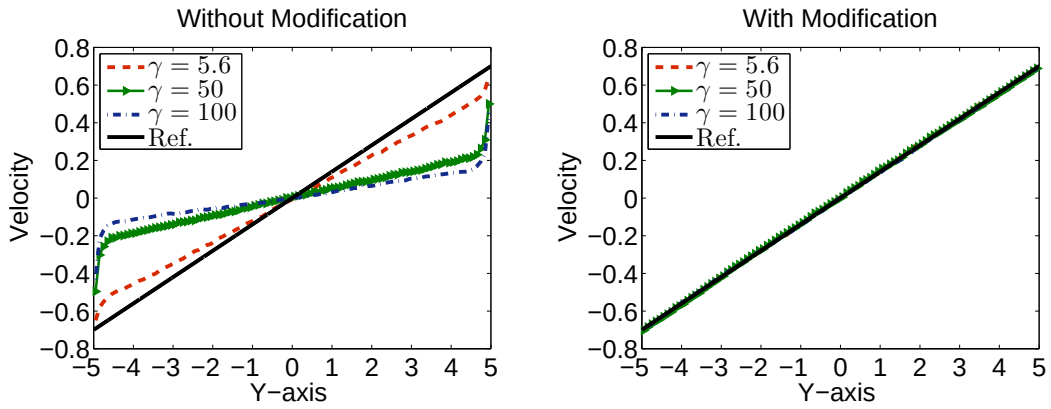


Figure 4.4: Comparisons of the computed velocity profile in dissipative particle dynamics (DPD) with different values of the friction coefficient γ under Lees–Edwards boundary conditions with shear rate $\dot{\gamma} = 0.14$ "without" (left) and "with" (right) suitable modification in the relative velocity, respectively. The black solid line is the expected linear profile. The DPD-S1 method was used with small stepsize $h = 0.01$.

boundaries where adjacent layers have different streaming velocities (i.e., the y-direction in Figure 2.2), was proposed in the original paper [30], followed by other suggestions [21, 82, 122].

However, the finding of [30] directly contradicts the principle of LEBC, which is translationally invariant and is intended to overcome the surface effects. In fact, as pointed in [51], in no way can the particles actually sense the boundaries of any given box since the system is spatially homogeneous. Furthermore, our numerical experiments, which is in perfect agreement with theoretically expected behavior as shown in the right panel of Figure 4.4 even in the high friction regime, suggest that the LEBC might have been incorrectly implemented in [30]. One possibility is that when calculating the relative velocity between one particle and an image of another, which is in a layer with different streaming velocity from its origin, the original velocity, rather than the "modified velocity" because of the different streaming velocities in different layers, was used as the velocity of the image particle. By neglecting the necessary modification, we obtained the left panel of Figure 4.4, while if the velocity of the image particle was modified properly the expected liner velocity profile was recovered in Figure 4.4 (right) using otherwise exactly the same setting.

Overall, it should be emphasized that if one particle is interacting with an image of another under certain conditions, the relative velocity (in the x-direction) between them should be modified as follows:

$$N_{\mathrm{L}} = 0 \tag{4.5}$$

$$\text{if } (\mathrm{fabs}(q_{ij}^{y}) > L_y/2) \quad N_{\mathrm{L}} = \mathrm{round}(q_{ij}^{y}/L_y) \tag{4.6}$$

$$\hat{v}_{ij}^{x} = v_{ij}^{x} - N_{\mathrm{L}}\dot{\gamma}L_y \tag{4.7}$$

where function "fabs($\cdot$)" returns the absolute value of the argument. Note that:

(1) q_{ij}^y in (4.6) has to be evaluated either before the minimum image convention (2.65)–(2.69) or by other proper ways to determine the actual distance between two interacting particles; (2) v_{ij}^x in (4.7) is the relative ("absolute") velocity between the two particles and one has to take into account the effects of the streaming velocity as indicated.

We highly suspect that the necessary modification (4.5)–(4.7) was not correctly implemented in [30], resulting in the unphysical behavior as shown in Figure 4.4. It is not surprising that by switching off the DPD thermostat on interactions that cross certain boundaries would recover the expected linear velocity profile as shown in [30], since it directly avoids the situation described in (4.5)–(4.7) where special attention should be paid. Overall, the "workaround" does not provide any physical explanation, and could affect the underlying dynamics of the system, implying that it should be abandoned.

Another interesting question in NEMD is that "What is the most appropriate way to measure the system temperature?" Since the streaming velocity should be subtracted from the particle velocity before calculating the kinetic temperature, the normal definition (3.73) should be modified as follows [51]:

$$k_{\mathrm{B}}\hat{T}_{\mathrm{k}} = \frac{1}{N_{\mathrm{d}}}\sum_i m_i(\mathbf{v}_i - \hat{\mathbf{u}})\cdot(\mathbf{v}_i - \hat{\mathbf{u}})\,, \tag{4.8}$$

where $\hat{\mathbf{u}}$ is the corresponding streaming velocity at the location of particle i.

If we can assume that the velocity profile is linear, as in uniform shear flow, we can just calculate and subsequently subtract it. However, as pointed in [51, 52], at higher shear rates and/or Reynolds numbers (i.e., the ratio of the inertial and viscous forces), the assumption of a linear streaming velocity profile is extremely dubious, even though LEBC are used. This issue was addressed in [51, 52] where the so-called profile-unbiased thermostat (PUT) was proposed. The PUT allows the simulation itself define the local streaming velocity (see more details in [51, 52, 176]). However, this is still not completely satisfactory since PUT assumes that the streaming velocity profile is stationary in time, whereas the profile could vary in time.

In DPD and related systems, temperature calculations can be based on relative velocities (e.g., (3.55) of NHLA), which do not rely on the relationship between the absolute particle velocity and an underlying streaming velocity. As discussed in Section 3.7.2, it might be even more desirable, especially in NEMD, to define the temperature solely based on the configurations, which leads to the "configurational temperature" (3.74) (see applications in [39, 40, 167]). In addition, thermostats based on the configurational temperature have been widely used in NEMD with shear flows [23, 41, 42, 115]. Therefore, the configurational temperature formulation will be preferred in our numerical experiments.

Figure 4.5 compares the configurational temperature control of various systems described under LEBC with different shear rates. As can be shown from the

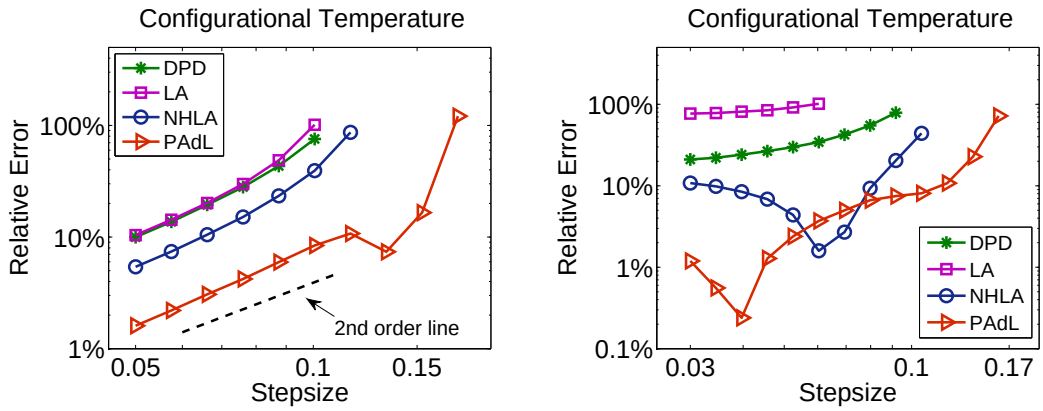


Figure 4.5: Comparisons of the relative error in computed configurational temperature against stepsize by using various numerical methods with (effective) friction coefficient $\gamma = 4.5$ under Lees–Edwards boundary conditions with shear rate $\dot{\gamma} = 0.2$ (left) and $\dot{\gamma} = 2$ (right). The format of the plots is the same as in Figure 4.1.

Method	Critical Stepsize	Maximal Stepsize	CPU Time	Scaled Efficiency
DPD-VV	0.05	0.10	20.212	100.0%
DPD-S1	0.05	0.11	20.618	98.0%
DPD-Trotter	0.05	0.11	21.451	94.2%
Peters	0.05	0.11	21.274	95.0%
LA	0.05	0.10	18.048	112.0%
NHLA	0.07	0.13	18.691	151.4%
PAdL	0.13	0.17	23.103	227.5%

Table 4.4: Comparisons of the computational efficiency of the various numerical methods in the moderate (effective) friction regime of $\gamma = 4.5$ under Lees–Edwards boundary conditions with shear rate $\dot{\gamma} = 0.2$. The format of the table is the same as in Table 4.2.

figure, when the shear rate is relatively small ($\dot{\gamma} = 0.2$, which is larger than that of Figure 4.4), the behavior is largely similar to that of Figure 4.1 (left): all the methods appear to show second order convergence to the invariant measure, and the newly proposed PAdL method achieves one order of magnitude improvement in numerical accuracy over both DPD and LA methods, both of which are slightly worse than the NHLA method. The overall numerical efficiency was compared in Table 4.4. Again the PAdL method is by far the most efficient method of all, which has an about 130% improvement over the benchmark DPD-VV method.

When the shear rate is relatively high ($\dot{\gamma} = 2$), as shown in Figure 4.5 (right), all the methods appear to lose the clear second order convergence previously observed. The LA thermostat, displaying significantly large relative error even when the stepsize is relatively small, appears to be most vulnerable to the high shear rate, with the DPD method being slightly better. While exhibiting some unexpected behavior, both NHLA and PAdL have better numerical accuracy control

than the other two. For instance, as the stepsize increases, the relative error of the PAdL method starts to decrease before growing up as would be expected. We believe this unexpected decrease at the beginning is due to sampling errors, which is more likely to be observed in high accuracy regime, since increasing the system size (i.e., number of particles) would resolve this issue. Nevertheless, the PAdL method constantly achieves at least one order of magnitude improvement in numerical accuracy over the DPD method.

Chapter 5

Noisy Gradient Systems with Applications in Data Science

5.1 Introduction

Stochastic thermostats [104, 150, 151] are powerful tools for sampling probability measures on high-dimensional spaces. These methods combine an extended dynamics with degenerate stochastic perturbation to ensure ergodicity. The traditional use of thermostats in molecular dynamics is to sample a well-specified equilibrium system involving a known force field which is the gradient of a potential energy function. Recently, however, these techniques have become increasingly popular for problems of more general form, including the following:

- multiscale models in which the forces are obtained by approximate sampling in another scale regime [54, 72, 109, 121, 140, 141];
- nonequilibrium physical models in which the potential energy function either is evolving or does not completely specify the system [87, 111, 133, 135, 158, 159];
- Bayesian machine learning applications in which a dataset defines an objective function which leads to an effective force law [4, 32, 44, 134, 173, 175].

In this chapter, we consider thermostats and numerical methods for sampling an underlying probability measure in the presence of error, under the assumption that the errors are random with a simple distributional form and unknown, but constant or slowly varying, parameters. In the cases considered, these methods are simple to implement, robust, and efficient. The results presented in this chapter are mostly from a recent work with B. Leimkuhler [107] and Section 5.5 is based on a recent article in collaboration with Z. Zhu, B. Leimkuhler, and A. Storkey [154].

5.1.1 Noisy Gradients

The gradient (or Hamiltonian) structure is essential to the nature of all the methods described in Chapter 2 since it is only by use of this feature that the underlying Fokker–Planck equation can be shown to have the desired steady state solution. However, in many applications, in particular multiscale modelling, the force is corrupted by significant approximation error and cannot be viewed as the gradient of a single global potential function. One imagines a large extended system involving configurational variables $\mathbf{q}$ and $\mathbf{y}$, with $(\mathbf{q}, \mathbf{y}) \in \Omega_{\mathbf{q}} \times \Omega_{\mathbf{y}}$ (compact), and an overall distribution described by a Gibbs–Boltzmann density

$$\tilde{\rho}(\mathbf{q}, \mathbf{y}) = Z^{-1} \exp\left(-\beta \tilde{U}(\mathbf{q}, \mathbf{y})\right) ,$$

where Z is a normalizing constant so that $\tilde{\rho}$ is a probability density. One calculates the mean force acting on $\mathbf{q}$, $\hat{f}(\mathbf{q})$, by averaging the forces in the extended Gibbsian system, $\tilde{f}(\mathbf{q}, \mathbf{y})$, as

$$\hat{f}(\mathbf{q}) = \int_{\Omega_{\mathbf{y}}} \tilde{f}(\mathbf{q}, \mathbf{y}) \tilde{\rho}(\mathbf{q}, \mathbf{y}) \, \mathrm{d}\omega_{\mathbf{y}} .$$

If, as would typically be assumed, $\tilde{f}(\mathbf{q}, \mathbf{y}) = -\nabla_{\mathbf{q}} \tilde{U}(\mathbf{q}, \mathbf{y})$, i.e., the force in the extended system is conservative, then we may interpret $\hat{f}$ as a conservative force as well, specifically the gradient of the potential of mean force, which is given by

$$\hat{U}(\mathbf{q}) = -\beta^{-1} \ln \int_{\Omega_{\mathbf{y}}} \exp\left(-\beta \tilde{U}(\mathbf{q}, \mathbf{y})\right) \, \mathrm{d}\omega_{\mathbf{y}} .$$

The challenge arises when this integral must be approximated. For example, if this is done by Monte Carlo integration, for fixed $\mathbf{q}$, one generates samples $\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^k$ from the distribution with density $\tilde{\rho}(\mathbf{q}, \mathbf{y})$ and thus approximates the mean force by

$$\bar{f}^k(\mathbf{q}) = k^{-1} \sum_{i=1}^{k} \tilde{f}\left(\mathbf{q}, \mathbf{y}^i\right) .$$

In practice most systems constructed in this way, for example, those arising in mixed quantum and classical molecular models [20], will admit very substantial errors in the forces; that is,

$$\bar{f}^k(\mathbf{q}) = \hat{f}(\mathbf{q}) + \Delta^k(\mathbf{q}) .$$

Depending on the method of computation, it may be reasonable to assume that the errors Δ^k are normally distributed with zero mean, which is justified by the central limit theorem [10], but the variance of the errors is generally not known and will be dependent on the location $\mathbf{q}$ where they are computed; thus we would expect

$$\Delta^k(\mathbf{q}) \sim \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}^k(\mathbf{q})\right) , \tag{5.1}$$

where $\mathbf{\Sigma}^k(\mathbf{q})$ is an unknown covariance matrix. It should be noted that the

assumption of the errors being Gaussian distributed is also often adopted in Bayesian inverse problems [36] and elsewhere.

The most straightforward approach to the problem is to first treat the estimation problem for $\mathbf{\Sigma}^k$ separately, by some means, and then to use this within a standard Brownian or Langevin dynamics algorithm. The difficulty is that this requires a high level of local accuracy in the calculations, which is likely to be burdensome and involve redundant computation. What we would prefer to do is to resolve the correct target distribution by a global calculation.

This problem has recently been encountered in the data science community, where it has attracted considerable attention [4,32,44,134,173,175]. To illustrate, we consider the problem of Bayesian sampling [24, 145], where one is interested in correctly drawing states from a posterior probability density defined as

$$\pi(\boldsymbol{\theta}|\mathbf{X}) \propto \pi(\mathbf{X}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\,, \tag{5.2}$$

where $\boldsymbol{\theta}$ is the parameter vector of interest, $\mathbf{X}$ represents the entire dataset, and, $\pi(\mathbf{X}|\boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta})$ represent the likelihood and prior distributions, respectively. In these applications, the distribution parameters are interpreted as the configuration variables ($\boldsymbol{\theta} \equiv \mathbf{q}$). We introduce a potential energy $U(\boldsymbol{\theta})$ by defining $\pi(\boldsymbol{\theta}|\mathbf{X}) \propto \exp(-\beta U(\boldsymbol{\theta}))$; thus taking the logarithm of (5.2) gives

$$U(\boldsymbol{\theta}) = -\log \pi(\mathbf{X}|\boldsymbol{\theta}) - \log \pi(\boldsymbol{\theta})\,. \tag{5.3}$$

Assuming the data are independent and identically distributed (i.i.d.), the logarithm of the likelihood distribution can then be calculated as

$$\log \pi(\mathbf{X}|\boldsymbol{\theta}) = \sum_{i=1}^{N} \log \pi(\mathbf{x}_i|\boldsymbol{\theta})\,, \tag{5.4}$$

where N is the size of the entire dataset.

However, in machine learning applications, one often finds that directly sampling with the entire large-scale dataset is computationally intractable. For instance, the Markov chain Monte Carlo (MCMC) method [118] requires the calculation of the acceptance probability and the creation of informed proposals based on the whole dataset, while the gradient is evaluated through the whole dataset in the hybrid Monte Carlo (HMC) method [24, 45, 77], again resulting in severe computational complexity.

In order to improve the efficiency of simulation, the so-called stochastic gradient Langevin dynamics (SGLD) was recently proposed [175] based on using a random (and much smaller, i.e., $\tilde{n} \ll N$) subset to approximate the likelihood of the dataset for given parameters,

$$\log \pi(\mathbf{X}|\boldsymbol{\theta}) \approx \frac{N}{\tilde{n}} \sum_{i=1}^{\tilde{n}} \log \pi(\mathbf{x}_{r_i}|\boldsymbol{\theta})\,, \tag{5.5}$$

where $\{\mathbf{x}_{r_i}\}_{i=1}^{\tilde{n}}$ represents a random subset of $\mathbf{X}$. Overall, the "noisy" potential energy now can be written as

$$\tilde{U}(\boldsymbol{\theta}) = -\frac{N}{\tilde{n}} \sum_{i=1}^{\tilde{n}} \log \pi(\mathbf{x}_{r_i}|\boldsymbol{\theta}) - \log \pi(\boldsymbol{\theta}) \,, \tag{5.6}$$

with "noisy" force $\tilde{\mathbf{F}}(\boldsymbol{\theta}) = -\nabla \tilde{U}(\boldsymbol{\theta})$.

5.1.2 Sampling Methods for Noisy Gradients

The challenge is to identify a method to compute samples distributed according to the Gibbs distribution $\rho(\mathbf{q}) = Z^{-1} \exp(-\beta U(\mathbf{q}))$, where the only available information is a stochastically perturbed force $\tilde{\mathbf{F}}(\mathbf{q})$ defined in the previous section.

In the original SGLD method, samples are generated by Brownian dynamics,

$$\mathbf{q}_{n+1} = \mathbf{q}_n + h_n \tilde{\mathbf{F}}(\mathbf{q}_n) + \sqrt{2\beta^{-1} h_n} \mathbf{R}_n \,, \tag{5.7}$$

where $\mathbf{R}_n$ is a vector of i.i.d. standard normal random variables. It should be emphasized that h_n is a sequence of stepsizes decreasing to zero [175]. Although a central limit theorem associated with the decreasing stepsize sequence was established by Teh et al. [165], a fixed stepsize is often preferred in practice, which is the choice in this thesis as in Vollmer et al. [173], where a modified SGLD (mSGLD) is introduced:

$$\mathbf{q}_{n+1} = \mathbf{q}_n + h \tilde{\mathbf{F}}(\mathbf{q}_n) + \sqrt{2\beta^{-1} h} \left(\mathbf{I} - \frac{h}{4} \mathrm{Cov} \tilde{\mathbf{F}}(\mathbf{q}_n) \right) \mathbf{R}_n \,, \tag{5.8}$$

where

$$\mathrm{Cov} \tilde{\mathbf{F}}_{ij} = \mathbb{E} \left[\left(\tilde{\mathbf{F}}_i - \mathbb{E}(\tilde{\mathbf{F}}_i) \right) \left(\tilde{\mathbf{F}}_j - \mathbb{E}(\tilde{\mathbf{F}}_j) \right)^T \right] \tag{5.9}$$

is the covariance matrix of the noisy force.

A stochastic gradient Hamiltonian Monte Carlo (SGHMC) method was also proposed recently by Chen et al. [32], which incorporates a parameter-dependent diffusion matrix $\boldsymbol{\Sigma}(\mathbf{q})$ (i.e., the covariance matrix of the noisy force, which can be calculated as in (5.9)). $\boldsymbol{\Sigma}(\mathbf{q})$ is intended to effectively offset the stochastic perturbation of the gradient. However, it is very difficult to accommodate $\boldsymbol{\Sigma}(\mathbf{q})$ in practice; moreover, as pointed out in [44], poor estimation of it may have a significant adverse influence in correctly sampling the target distribution unless the stepsize is small enough.

These problems challenge the conventional mechanism of thermostats. An article of Jones and Leimkuhler [84] provides an alternative means of tackling this problem by showing that Nosé–Hoover dynamics is able to adaptively dissipate excess heat pumped into the system while maintaining the Gibbs (canonical) dis-

tribution. In the setting of systems involving a driving stochastic perturbation, the adaptive Nosé–Hoover method is referred to as Ad-NH, with similar generalizations of Nosé–Hoover–Langevin (Ad-NHL) and Langevin dynamics (Ad-Langevin) available. An idea equivalent to Ad-Langevin was recently applied in the setting of Bayesian sampling for use in data science calculations by Ding et al. [44], which they referred to as the stochastic gradient Nosé–Hoover thermostat (SGNHT). It showed significant advantages over alternative techniques such as SGHMC [32]. However, the numerical method used by Ding et al. [44] is not optimal, neither in terms of its accuracy (measured per unit work) nor its stability (measured by the largest usable stepsize).

Although extended systems have been increasingly popular in molecular simulations, the mathematical analysis of the order of convergence, specifically in terms of the bias in averaged quantities computed using numerical trajectories, is not fully understood. Using a splitting approach, we propose in this section an alternative numerical method for Ad-Langevin simulation that substantially improves on the existing schemes in the literature in terms of accuracy, robustness, and overall numerical efficiency.

5.2 Adaptive Thermostats for Noisy Gradients

In this section, we discuss the construction of thermostats to approximate samples with respect to the target measure (i.e., the correct marginalized Gibbs density) if the covariance matrix of the noisy force is constant, i.e., $\mathbf{\Sigma}(\mathbf{q}) = \sigma^2\mathbf{I}$ (σ is a constant positive quantity). The procedure was outlined in the paper of Jones and Leimkuhler [84] and relies on the fact that a fixed amplitude noise perturbation engenders a shift of the auxiliary variable in the extended stationary distribution associated with the Nosé–Hoover thermostat.

If the system is not coming from a Newtonian dynamics model, then it is unclear that we need to rely on second order dynamics for this purpose. To see why this is the case, we explain what goes wrong if we try to use first order dynamics. In what follows, we assume that the covariance matrix of the noisy force is constant, although we ultimately intend to apply the method more generally. Even in the constant σ case it is a nontrivial problem to extract statistics related to a particular target temperature, since we do not assume that σ is known.

For σ constant, let us first consider the SDE

$$\mathrm{d}\mathbf{q} = -\xi\nabla U(\mathbf{q})\mathrm{d}t + \sigma\mathrm{d}\mathbf{W}\,, \tag{5.10}$$

$$\mathrm{d}\xi = \chi(\mathbf{q})\mathrm{d}t \tag{5.11}$$

and seek $\chi(\cdot)$ so that an extended Gibbs distribution with density of the form $\psi(\mathbf{q},\xi) = \bar{\rho}_\beta(\mathbf{q})\varphi(\xi)$ is (ergodically) preserved. The variable ξ is an auxiliary variable. We do not generally care what its distribution is, but it is crucial that

(i) the overall density is in product form, and

(ii) $\varphi(\xi) \geq 0$ is normalizable and of a simple, easily sampled form.

These conditions ensure that we can easily average out over the auxiliary variable to compute the averages of functions of $\mathbf{q}$ which are of greatest interest.

Proposition 1. *Let* $\chi(\mathbf{q}) = -\beta^{-1}\Delta U(\mathbf{q}) + \|\nabla U(\mathbf{q})\|^2$*; then* (5.10)–(5.11) *preserves the modified Gibbs distribution:*

$$\tilde{\rho}(\mathbf{q}, \xi) = \bar{\rho}_\beta(\mathbf{q}) e^{-\beta(\xi-\hat{\gamma})^2/2}\,,$$

where $\hat{\gamma} = \beta\sigma^2/2$*.*

Proof. The Fokker–Planck equation corresponding to (5.10)–(5.11) is

$$\rho_t = \mathcal{L}^\dagger \rho := \xi \nabla \cdot (\nabla U(\mathbf{q})\rho(\mathbf{q}, \xi)) + \frac{\sigma^2}{2}\Delta\rho - \frac{\partial}{\partial \xi}(\chi(\mathbf{q})\rho)\,.$$

Just insert $\tilde{\rho}$ into the operator $\mathcal{L}^\dagger$ to see that it vanishes. □

Proposition 1 tells us that if we can solve system (5.10)–(5.11), under an assumption of ergodicity, we can compute averages with respect to the target Gibbs distribution without actually knowing the value of σ. σ could be observed retrospectively by simply averaging ξ during simulation, since $\langle \xi \rangle = \beta\sigma^2/2$.

The problem is that the dynamics (5.10)–(5.11) is not quite what we want. A typical numerical method for this system might be constructed based on modification of the Euler–Maruyama method:

$$\mathbf{q}_{n+1} = \mathbf{q}_n - h\xi_n \nabla U(\mathbf{q}_n) + \sigma\sqrt{h}\mathbf{R}_n\,, \tag{5.12}$$

$$\xi_{n+1} = \xi_n + h\chi(\mathbf{q}_n)\,; \tag{5.13}$$

however, observe that this method requires separate knowledge of $\nabla U(\mathbf{q})$ and σ, which is generally impossible a priori, as we assume that the force is polluted by unknown noise. The form of the equations means that we evaluate the product of ξ and the deterministic force, on the one hand, and the random perturbation, on the other hand, separately, and these contributions are independently scaled by h and $\sqrt{h}$, respectively.

5.2.1 Adaptive Langevin (Ad-Langevin) Thermostat

To adaptively control the invariant distribution, we consider the following second order formulation, which was first introduced in the paper of Jones and Leimkuh-

ler [84]:

$$\begin{aligned} \mathrm{d}\mathbf{q} &= \mathbf{M}^{-1}\mathbf{p}\mathrm{d}t\,, \\ \mathrm{d}\mathbf{p} &= \tilde{\mathbf{F}}(\mathbf{q})\mathrm{d}t - \xi\mathbf{p}\mathrm{d}t + \sigma_{\mathrm{A}}\mathbf{M}^{1/2}\mathrm{d}\mathbf{W}_{\mathrm{A}}\,, \\ \mathrm{d}\xi &= \mu^{-1}\left[\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p} - N_{\mathrm{d}}k_{\mathrm{B}}T\right]\mathrm{d}t\,. \end{aligned} \tag{5.14}$$

In these equations, $\tilde{\mathbf{F}}(\mathbf{q})$ is meant to represent a noisy gradient which may be thought of as being defined by the relation

$$\tilde{\mathbf{F}}(\mathbf{q}) = -\nabla U(\mathbf{q}) + \sigma\mathbf{M}^{1/2}\mathbf{R}\,, \tag{5.15}$$

where $\mathbf{R} = \mathbf{R}(t)$ is a collection of independent Gaussian white noise processes, i.e., $\langle \mathrm{R}_i(t)\mathrm{R}_j(s)\rangle = \delta_{ij}\delta(t-s)$. $\sigma_{\mathrm{A}}\mathbf{M}^{1/2}\mathrm{d}\mathbf{W}_{\mathrm{A}}$ indicates the artificial noise added into the system to enhance the ergodicity; i.e., the constant σ_{A} is known a priori. All the components of the Wiener process $\mathbf{W}_{\mathrm{A}}(t)$ are assumed to be independent. N_{d} denotes the number of degrees of freedom of the system. μ is a coupling parameter which is referred to as the "thermal mass". k_{B} and T, satisfying the relation $\beta^{-1} = k_{\mathrm{B}}T$, represent the Boltzmann constant and system temperature, respectively.

A similar system (SGNHT) was used by Ding et al. [44], who also explored its application to three examples from machine learning. These experiments demonstrated that Ad-Langevin has superior performance compared to SGHMC in various applications, confirming the importance of adaptively dissipating additional noise in sampling. However, there remain two important issues that we wish to address in this chapter: (1) the underlying dynamics of the Ad-Langevin method is not clear due to the presence of the stochastically perturbed gradient; (2) little attention has been paid to the design of optimal numerical methods for implementing Ad-Langevin with attention to stability and numerical efficiency.

One may wonder why the artificial noise is needed (i.e., $\sigma_{\mathrm{A}} \neq 0$), since we are assuming the presence of noise in the gradient itself. The reason is as follows: in defining a numerical method for the noisy gradient system, the force (including the random perturbation) will in general be multiplied by h, where h is the timestep. On the other hand, the Itō rule implies that the scaling of random perturbations in an SDE should be by a factor proportional to $\sqrt{h}$; thus, effectively, if we are to relate the thermostatted method to a standard SDE, the standard deviation of the noise is reduced by multiplication by the factor $\sqrt{h}$. The noise perturbation introduced at each timestep (and the effective diffusion) is thus reduced for small stepsizes and it is therefore important to inject additional artificial noise in order to stabilize the invariant distribution. A rewriting of the Ad-Langevin system as a standard Itō SDE system makes clear the relation between the different terms

$$\begin{aligned} \mathrm{d}\mathbf{q} &= \mathbf{M}^{-1}\mathbf{p}\mathrm{d}t\,, \\ \mathrm{d}\mathbf{p} &= -\nabla U(\mathbf{q})\mathrm{d}t + \sigma\sqrt{h}\mathbf{M}^{1/2}\mathrm{d}\mathbf{W} - \xi\mathbf{p}\mathrm{d}t + \sigma_{\mathrm{A}}\mathbf{M}^{1/2}\mathrm{d}\mathbf{W}_{\mathrm{A}}\,, \\ \mathrm{d}\xi &= \mu^{-1}\left[\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p} - N_{\mathrm{d}}k_{\mathrm{B}}T\right]\mathrm{d}t\,, \end{aligned} \tag{5.16}$$

where $\mathbf{W} = \mathbf{W}(t)$ is an additional vector of standard Wiener processes.

Let us note the main features of the dynamics (5.16):

(i) The equations are a combination of Langevin dynamics and Nosé–Hoover dynamics. If ξ is constant in the equation for the momentum, then the system reduces to Langevin dynamics. In the absence of noise, $\sigma_{\mathrm{A}} = 0$ (and $\sigma = 0$); then the system reduces to Nosé–Hoover. The system (5.16) may be regarded as a sort of Langevin dynamics where the friction coefficient, rather than being fixed a priori, is automatically and adaptively determined in order to achieve the desired temperature (which is specified in the control law defining the evolution of ξ).

(ii) The invariant distribution for the given system may be directly obtained by study of its Fokker–Planck equation. Following [84], it is straightforward to show that (5.16) has the following invariant distribution:

$$\tilde{\rho}_\beta(\mathbf{q}, \mathbf{p}, \xi) = \frac{1}{Z} \exp\left(-\beta H(\mathbf{q}, \mathbf{p})\right) \exp\left(-\frac{\beta\mu}{2}(\xi - \hat{\gamma})^2\right), \tag{5.17}$$

where Z is the normalizing constant and

$$\hat{\gamma} = \frac{\beta\left(\sigma_{\mathrm{F}}^2 + \sigma_{\mathrm{A}}^2\right)}{2}, \tag{5.18}$$

where $\sigma_{\mathrm{F}} = \sigma\sqrt{h}$. Observe that this means that if $\sigma_{\mathrm{A}} = 0$, then, as $\lim_{h\to 0}\sigma_{\mathrm{F}} = 0$, we find that ξ tends to a variable which is normally distributed with mean zero. Alternatively, if $\sigma_{\mathrm{A}} \neq 0$, one would obtain

$$\xi \xrightarrow{\mathscr{L}} \mathcal{N}\left(\frac{\beta\sigma_{\mathrm{A}}^2}{2}, \beta^{-1}\mu^{-1}\right), \quad t \to \infty, \quad h \to 0,$$

where $\beta^{-1}\mu^{-1}$ is the variance and the symbol $\xrightarrow{\mathscr{L}}$ indicates that ξ converges in probability law to a normally distributed random variable with the indicated parameters. The order of the limits here is important: $t \to \infty$ first (to reach the invariant distribution), then $h \to 0$.

(iii) The ergodicity of (5.16) with respect to the distribution indicated above can easily be demonstrated by reference to Hörmander's condition for hypoellipticity following the method in [116], as for Langevin dynamics. The only additional step is to verify that the noise propagates into the ξ variable, which follows due to its strong coupling to the momenta.

(iv) This dynamics is a bit unusual in that it must be viewed as stepsize dependent, although we mention that such mixed systems are used in the study of backward error analysis [105]. One simply thinks of the characteristics of stochastic paths associated with (5.16) as being stepsize dependent. Although (5.16) takes on the appearance of a standard Itō SDE system, we must bear in mind that in discretizing these equations the conservative

force $\mathbf{F}(\mathbf{q})$ and the associated noise term $\sigma\sqrt{h}\mathbf{M}^{1/2}\mathrm{d}\mathbf{W}$ must be evaluated together at every stage, since the formulation (5.16) is a notational device to make clear the properties of the system.

5.3 Numerical Methods for Adaptive Thermostats

Since stochastic systems in most of the cases cannot be solved "exactly", splitting methods are often adopted in practice. For instance here, the vector field of the Ad-Langevin/SGNHT (5.14) can be split into four pieces which are denoted as "A", "B", "O", and "D", in such a way that each piece can be solved "exactly",

$$\mathrm{d}\begin{bmatrix}\mathbf{q}\\ \mathbf{p}\\ \xi\end{bmatrix} = \underbrace{\begin{bmatrix}\mathbf{M}^{-1}\mathbf{p}\\ \mathbf{0}\\ 0\end{bmatrix}\mathrm{d}t}_{\mathrm{A}} + \underbrace{\begin{bmatrix}\mathbf{0}\\ -\nabla U(\mathbf{q})+\sigma\mathbf{M}^{1/2}\mathbf{R}\\ 0\end{bmatrix}\mathrm{d}t}_{\mathrm{B}} + \underbrace{\begin{bmatrix}\mathbf{0}\\ -\xi\mathbf{p}\mathrm{d}t+\sigma_{\mathrm{A}}\mathbf{M}^{1/2}\mathrm{d}\mathbf{W}_{\mathrm{A}}\\ 0\end{bmatrix}}_{\mathrm{O}} + \underbrace{\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\\ G(\mathbf{p})\end{bmatrix}\mathrm{d}t}_{\mathrm{D}},$$

where $G(\mathbf{p}) = \mu^{-1}\left[\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p} - N_{\mathrm{d}}k_{\mathrm{B}}T\right]$.

Clearly parts "A" and "D" can be solved "exactly". As mentioned previously, the underlying dynamics for "B" is

$$\mathrm{d}\mathbf{p} = -\nabla U(\mathbf{q})\mathrm{d}t + \sigma_{\mathrm{F}}\mathbf{M}^{1/2}\mathrm{d}\mathbf{W}\,, \tag{5.19}$$

where $\mathbf{q}$ is fixed and $\sigma_{\mathrm{F}} = \sigma\sqrt{h}$. Integrating (5.19) from 0 to h gives the exact solution in distribution of this part as

$$\begin{aligned}\mathbf{p}(h) &= \mathbf{p}(0) - h\nabla U(\mathbf{q}) + \sqrt{h}\sigma_{\mathrm{F}}\mathbf{M}^{1/2}\mathbf{R}\\ &= \mathbf{p}(0) + h[-\nabla U(\mathbf{q}) + \sigma\mathbf{M}^{1/2}\mathbf{R}] = \mathbf{p}(0) + h\tilde{\mathbf{F}}(\mathbf{q})\,,\end{aligned}$$

where $\mathbf{R}$ is a vector of i.i.d. standard normal random variables. It should be noted that applying the Euler–Maruyama method to (5.19) gives the same result; thus, for constant force, Euler–Maruyama is "exact".

The "O" or "Ornstein–Uhlenbeck" part is usually stated with ξ a positive constant, in which case the solution is found to be [90] (see also (3.20))

$$\mathbf{p}(h) = e^{-\xi h}\mathbf{p}(0) + \sigma_{\mathrm{A}}\sqrt{\frac{1-e^{-2\xi h}}{2\xi}}\mathbf{M}^{1/2}\mathbf{R}\,, \tag{5.20}$$

where $\mathbf{p}(0)$ is the initial value of the variable and $\mathbf{R}$ is a vector of i.i.d. standard normal random variables. However, the same formula (5.20) is easily seen to be

valid for $\xi < 0$, since the quantity $(1 - e^{-2\xi h})/(2\xi)$ is strictly greater than zero unless $\xi = 0$. (The proof is obtained by following the standard procedure [90].) When $\xi = 0$, one can simply replace $(1-e^{-2\xi h})/(2\xi)$ by its well-defined asymptotic limit,

$$\mathbf{p}(h) = \mathbf{p}(0) + \sqrt{h}\sigma_{\mathrm{A}}\mathbf{M}^{1/2}\mathbf{R}\,. \tag{5.21}$$

The generators associated with each piece are defined, respectively, as

$$\begin{aligned}
\mathcal{L}_{\mathrm{A}} &= \mathbf{M}^{-1}\mathbf{p}\cdot\nabla_{\mathbf{q}}\,,\\
\mathcal{L}_{\mathrm{B}} &= -\nabla U(\mathbf{q})\cdot\nabla_{\mathbf{p}} + \frac{\sigma_{\mathrm{F}}^2}{2}\mathrm{Tr}\left(\mathbf{M}\nabla_{\mathbf{p}}^2\right)\,,\\
\mathcal{L}_{\mathrm{O}} &= -\xi\mathbf{p}\cdot\nabla_{\mathbf{p}} + \frac{\sigma_{\mathrm{A}}^2}{2}\mathrm{Tr}\left(\mathbf{M}\nabla_{\mathbf{p}}^2\right)\,,\\
\mathcal{L}_{\mathrm{D}} &= G(\mathbf{p})\frac{\partial}{\partial\xi}\,,
\end{aligned}$$

where $\sigma_{\mathrm{F}} = \sigma\sqrt{h}$ in part "B" is stepsize dependent.

Overall, the generator of the Ad-Langevin/SGNHT (5.14) system can be written as

$$\mathcal{L} = \mathcal{L}_{\mathrm{A}} + \mathcal{L}_{\mathrm{B}} + \mathcal{L}_{\mathrm{O}} + \mathcal{L}_{\mathrm{D}}\,. \tag{5.22}$$

Again, the flow map of the system, $\mathcal{F}_t$, can be written in the shorthand notation and approximations of it can be obtained as products (taken in different arrangements) of exponentials of the splitting terms. For example, the phase space propagation of the method proposed by Ding et al. [44] for the Ad-Langevin/SGNHT (5.14) system (denoted as "SGNHT-N") can be written as

$$\exp\left(h\hat{\mathcal{L}}_{\mathrm{SGNHT-N}}\right) = \exp\left(h\mathcal{L}_{\mathrm{P}}\right)\exp\left(h\mathcal{L}_{\mathrm{A}}\right)\exp\left(h\mathcal{L}_{\mathrm{D}}\right)\,, \tag{5.23}$$

where

$$\mathcal{L}_{\mathrm{P}} = \mathcal{L}_{\mathrm{B}} + \mathcal{L}_{\mathrm{O}} \tag{5.24}$$

and $\exp\left(h\mathcal{L}_f\right)$ represents the phase space propagator associated with the corresponding vector field f. Because of its nonsymmetric structure, one anticipates first order convergence to the invariant measure (for any choice of σ). Due to the naming of the component parts, the SGNHT-N method may be denoted by "PAD".

Overall, the SGNHT-N/PAD integration method is as follows:

$$\begin{aligned}
\mathbf{p}_{n+1} &= \mathbf{p}_n + h\left(-\nabla U(\mathbf{q}_n) + \sigma\mathbf{M}^{1/2}\mathbf{R}'_n\right) - h\xi_n\mathbf{p}_n + \sqrt{h}\sigma_{\mathrm{A}}\mathbf{M}^{1/2}\mathbf{R}_n\,,\\
\mathbf{q}_{n+1} &= \mathbf{q}_n + h\mathbf{M}^{-1}\mathbf{p}_{n+1}\,,\\
\xi_{n+1} &= \xi_n + h\mu^{-1}\left(\mathbf{p}_{n+1}^T\mathbf{M}^{-1}\mathbf{p}_{n+1} - N_{\mathrm{d}}k_{\mathrm{B}}T\right)\,,
\end{aligned}$$

where $\mathbf{R}'_n$ and $\mathbf{R}_n$ are vectors of i.i.d. standard normal random variables.

We propose symmetric alternative methods, such as the following symmetric Ad-Langevin/SGNHT (SGNHT-S) splitting method

$$e^{h\hat{\mathcal{L}}_{\mathrm{SGNHT-S}}} = e^{\frac{h}{2}\mathcal{L}_{\mathrm{B}}} e^{\frac{h}{2}\mathcal{L}_{\mathrm{A}}} e^{\frac{h}{2}\mathcal{L}_{\mathrm{D}}} e^{h\mathcal{L}_{\mathrm{O}}} e^{\frac{h}{2}\mathcal{L}_{\mathrm{D}}} e^{\frac{h}{2}\mathcal{L}_{\mathrm{A}}} e^{\frac{h}{2}\mathcal{L}_{\mathrm{B}}} , \tag{5.25}$$

where exact solvers for parts "B" and "O" derived above are applied. The SGNHT-S method may be referred to as "BADODAB", where it should be noted that the various operations are symmetrically applied and the steplengths are uniform and span the interval h. Other symmetric splittings are considered below.

The SGNHT-S numerical integration method may be written as

$$\begin{aligned}
\mathbf{p}_{n+1/3} &= \mathbf{p}_n + (h/2)\left(-\nabla U(\mathbf{q}_n) + \sigma \mathbf{M}^{1/2}\mathbf{R}'_n\right) , \\
\mathbf{q}_{n+1/2} &= \mathbf{q}_n + (h/2)\mathbf{M}^{-1}\mathbf{p}_{n+1/3} , \\
\xi_{n+1/2} &= \xi_n + (h/2)\mu^{-1}\left(\mathbf{p}_{n+1/3}^T \mathbf{M}^{-1}\mathbf{p}_{n+1/3} - N_{\mathrm{d}} k_{\mathrm{B}} T\right) , \\
\text{if } (\xi_{n+1/2} \neq 0): \quad \mathbf{p}_{n+2/3} &= e^{-\xi_{n+1/2}h}\mathbf{p}_{n+1/3} + \sigma_{\mathrm{A}}\sqrt{(1-e^{-2\xi_{n+1/2}h})/(2\xi_{n+1/2})}\mathbf{M}^{1/2}\mathbf{R}_n , \\
\text{else}: \quad \mathbf{p}_{n+2/3} &= \mathbf{p}_{n+1/3} + \sqrt{h}\sigma_{\mathrm{A}}\mathbf{M}^{1/2}\mathbf{R}_n , \\
\xi_{n+1} &= \xi_{n+1/2} + (h/2)\mu^{-1}\left(\mathbf{p}_{n+2/3}^T \mathbf{M}^{-1}\mathbf{p}_{n+2/3} - N_{\mathrm{d}} k_{\mathrm{B}} T\right) , \\
\mathbf{q}_{n+1} &= \mathbf{q}_{n+1/2} + (h/2)\mathbf{M}^{-1}\mathbf{p}_{n+2/3} , \\
\mathbf{p}_{n+1} &= \mathbf{p}_{n+2/3} + (h/2)\left(-\nabla U(\mathbf{q}_{n+1}) + \sigma \mathbf{M}^{1/2}\mathbf{R}'_{n+1}\right) .
\end{aligned}$$

The force computed at the end of each timestep can be reused at the start of the next step; thus only one force calculation is needed in SGNHT-S at each timestep, the same as for SGNHT-N. In practice, one could replace the exponential and square root operations in the exact solver of the "O" part by their respective well-defined asymptotic expansions to reduce the computational cost.

5.3.1 Order of Convergence of Ad-Langevin/SGNHT

Following the procedure in Section 3.6.1, we can work out the leading operator $\mathcal{L}_1^\dagger$ associated with the nonsymmetric SGNHT-N/PAD method (5.23) of Ding et al. [44],

$$\mathcal{L}_{1,\mathrm{PAD}}^\dagger = \frac{1}{2}\left(\left[\mathcal{L}_{\mathrm{D}}^\dagger, \mathcal{L}_{\mathrm{A}}^\dagger\right] + \left[\mathcal{L}_{\mathrm{D}}^\dagger, \mathcal{L}_{\mathrm{P}}^\dagger\right] + \left[\mathcal{L}_{\mathrm{A}}^\dagger, \mathcal{L}_{\mathrm{P}}^\dagger\right]\right) . \tag{5.26}$$

It is clear that the leading term $f_{1,\mathrm{PAD}}$ in the perturbed distribution (3.65) is in general nonzero. Therefore the nonsymmetric SGNHT-N/PAD method would be expected to exhibit first order convergence to the invariant measure. It should be noted that if certain conditions are satisfied, higher order convergence to the invariant measure would be possible as demonstrated by Abdulle et al. [1, 2]. However, it can be easily demonstrated that it is not the case here for the SGNHT-N/PAD method. In the presence of a noisy gradient, the Ad-Langevin/SGNHT methods, despite the stepsize dependency (5.16), would sim-

ilarly (and generally) be expected to be first order with respect to the invariant distribution.

5.3.2 Superconvergence Property

Recently, it has been demonstrated in the setting of Langevin dynamics that a particular symmetric splitting method ("BAOAB"), which requires only one force calculation per step, is fourth order for configurational quantities in the ergodic limit and in the limit of large friction [99, 102].

In what follows we demonstrate that the newly proposed SGNHT-S/BADODAB method (5.25) effectively inherits the superconvergence property of BAOAB in the setting of Ad-Langevin/SGNHT system (5.16) with a clean gradient, in case where the parameters σ_{A} and μ are both taken to infinity in a suitable way. For simplicity, we consider here a one-dimensional model $H = p^2/2 + U(q)$, but the analysis could easily be extended to higher dimensions.

Following the standard procedure described in Section 5.3.1, we obtain the following PDE associated with the BADODAB method:

$$\mathcal{L}^\dagger(\tilde{\rho}_\beta f_2) = -\mathcal{L}_2^\dagger \tilde{\rho}_\beta \,, \tag{5.27}$$

where $\mathcal{L}^\dagger$ is the exact Fokker–Planck operator

$$\mathcal{L}^\dagger = -p\partial_q + U'(q)\partial_p + \xi\partial_p(p\cdot) + \frac{\hat{\gamma}}{\beta}\partial_{pp} - \frac{1}{\mu}(p^2 - \beta^{-1})\partial_\xi \tag{5.28}$$

with invariant measure

$$\tilde{\rho}_\beta(q,p,\xi) = \frac{1}{Z}\exp\left(-\beta H(q,p)\right)\exp\left(-\frac{\beta\mu}{2}(\xi-\hat{\gamma})^2\right) , \tag{5.29}$$

where $\hat{\gamma} = \langle\xi\rangle = \beta\sigma_{\mathrm{A}}^2/2$ and $\mathcal{L}_2^\dagger$ can be calculated by using the BCH expansion

$$\begin{aligned}
\mathcal{L}_2^\dagger = &\frac{1}{12}\left(\left[\mathcal{L}_{\mathrm{O}}^\dagger,\left[\mathcal{L}_{\mathrm{O}}^\dagger,\mathcal{L}_{\mathrm{D}}^\dagger\right]\right] + \left[\mathcal{L}_{\mathrm{D}}^\dagger + \mathcal{L}_{\mathrm{O}}^\dagger,\left[\mathcal{L}_{\mathrm{D}}^\dagger + \mathcal{L}_{\mathrm{O}}^\dagger,\mathcal{L}_{\mathrm{A}}^\dagger\right]\right]\right) \\
&+ \frac{1}{12}\left(\left[\mathcal{L}_{\mathrm{A}}^\dagger + \mathcal{L}_{\mathrm{D}}^\dagger + \mathcal{L}_{\mathrm{O}}^\dagger,\left[\mathcal{L}_{\mathrm{A}}^\dagger + \mathcal{L}_{\mathrm{D}}^\dagger + \mathcal{L}_{\mathrm{O}}^\dagger,\mathcal{L}_{\mathrm{B}}^\dagger\right]\right]\right) \\
&- \frac{1}{24}\left(\left[\mathcal{L}_{\mathrm{D}}^\dagger,\left[\mathcal{L}_{\mathrm{D}}^\dagger,\mathcal{L}_{\mathrm{O}}^\dagger\right]\right] + \left[\mathcal{L}_{\mathrm{A}}^\dagger,\left[\mathcal{L}_{\mathrm{A}}^\dagger,\mathcal{L}_{\mathrm{D}}^\dagger + \mathcal{L}_{\mathrm{O}}^\dagger\right]\right] + \left[\mathcal{L}_{\mathrm{B}}^\dagger,\left[\mathcal{L}_{\mathrm{B}}^\dagger,\mathcal{L}_{\mathrm{A}}^\dagger + \mathcal{L}_{\mathrm{D}}^\dagger + \mathcal{L}_{\mathrm{O}}^\dagger\right]\right]\right) ,
\end{aligned}$$

whose action on the extended invariant measure reads as

$$\begin{aligned}\mathcal{L}_2^\dagger \tilde{\rho}_\beta = &\frac{1}{12}\left[-\beta p^3 U'''(q) + 4\beta p^2 \xi^3 + 3\beta \xi p^2 U''(q) + 3\beta p U'(q) U''(q) + \frac{6\xi p^2}{\mu}\left(1 - \beta p^2\right)\right]\tilde{\rho}_\beta \\ &+ \frac{\hat{\gamma}}{12}\left[3U''(q) + 4\xi^2 - 16\beta p^2 \xi^2 - 6\beta U''(q) p^2 + \frac{6}{\mu}\left(2\beta p^4 - 5p^2 + \beta^{-1}\right)\right]\tilde{\rho}_\beta \\ &+ \hat{\gamma}^2 \xi \left(2\beta p^2 - 1\right)\tilde{\rho}_\beta + \hat{\gamma}^3\left(\frac{2}{3} - \beta p^2\right)\tilde{\rho}_\beta \,.\end{aligned}$$

The equation is very complicated and we have no direct means of solving it. However, the additional variable ξ has mean $\hat{\gamma}$. If we suppose that μ is large, then the variance of ξ will be small. In this case we can consider the approximation obtained by replacing functions of ξ in the PDE (5.27) by their corresponding averages

$$\langle \xi \rangle = \hat{\gamma}\,, \qquad \langle \xi^2 \rangle = \frac{1}{\beta\mu} + \hat{\gamma}^2\,, \qquad \langle \xi^3 \rangle = \frac{3\hat{\gamma}}{\beta\mu} + \hat{\gamma}^3\,. \tag{5.30}$$

We use this as part of an averaging of the stationary Fokker–Planck equation with respect to the auxiliary variable. That is, we project the Fokker–Planck equation and its solution by integrating with respect to the Gaussian distribution of ξ in the ergodic limit. We can think of this is as defining a sort of "subspace projection"; it is related to the Galerkin method that is widely used in solving high-dimensional linear systems and PDEs, including Fokker–Planck equations [29, 143]. In this case, we apply the projection operator [61]

$$\mathcal{P}\nu(q,p,\xi) := \frac{\int_{\Omega_\xi} \tilde{\rho}_\beta(q,p,\xi)\nu(q,p,\xi)\,\mathrm{d}\xi}{\int_{\Omega_\xi} \tilde{\rho}_\beta(q,p,\xi)\,\mathrm{d}\xi}\,, \tag{5.31}$$

where ν is an arbitrary function, to the PDE (5.27). Effectively, this results in the reduced equation

$$\check{\mathcal{L}}^\dagger(\rho_\beta \hat{f}_2) = -\rho_\beta \mathcal{P}\frac{\mathcal{L}_2^\dagger \tilde{\rho}_\beta}{\tilde{\rho}_\beta}\,, \tag{5.32}$$

where the operator $\check{\mathcal{L}}^\dagger$ is just the operator $\mathcal{L}^\dagger$ reduced by the action of the projection, and which acts on functions of q and p; this is nothing other than the corresponding adjoint generator of Langevin dynamics. Likewise, $\hat{f}_2$ is now a function of q and p only. The right-hand side simplifies to

$$\begin{aligned}\rho_\beta \mathcal{P}\frac{\mathcal{L}_2^\dagger \tilde{\rho}_\beta}{\tilde{\rho}_\beta} = &\left(\frac{\beta}{12}\left[3pU'(q)U''(q) - p^3 U'''(q)\right]\right)\rho_\beta \\ &+ \left(\frac{\hat{\gamma}}{12}\left[3U''(q) - 3\beta p^2 U''(q) + \frac{1}{\mu}\left(6\beta p^4 - 28p^2 + 10\beta^{-1}\right)\right]\right)\rho_\beta\,,\end{aligned}$$

where ρ_β is the Gibbs (canonical) density ($\exp(-\beta H(q,p))$).

We consider the high friction limit ($\hat{\gamma} \to \infty$) and expand $\hat{f}_2$ in a series involving

the reciprocal friction $\varepsilon = 1/\hat{\gamma}$,

$$\hat{f}_2(q,p) = \hat{f}_{2,0}(q,p) + \varepsilon \hat{f}_{2,1}(q,p) + \varepsilon^2 \hat{f}_{2,2}(q,p) + \cdots , \tag{5.33}$$

with each function $\hat{f}_{2,i}$ satisfying $\langle \hat{f}_{2,i} \rangle = 0$. Dividing (5.27) by the friction coefficient $\hat{\gamma}$, we obtain

$$\left(\bar{\mathcal{L}}_{\mathrm{O}}^{\dagger} + \varepsilon \mathcal{L}_{\mathrm{H}}^{\dagger} \right) \left(\hat{f}_{2,0} + \varepsilon \hat{f}_{2,1} + O(\varepsilon^2) \right) \rho_\beta = -\varepsilon \rho_\beta \mathcal{P} \frac{\mathcal{L}_2^{\dagger} \tilde{\rho}_\beta}{\tilde{\rho}_\beta} , \tag{5.34}$$

where

$$\bar{\mathcal{L}}_{\mathrm{O}}^{\dagger} = \partial_p (p \cdot) + \beta^{-1} \partial_{pp} , \qquad \mathcal{L}_{\mathrm{H}}^{\dagger} = -p \partial_q + U'(q) \partial_p . \tag{5.35}$$

We take the high thermal mass limit ($\mu \to \infty$) in such a way that $\varepsilon = 1/\mu = 1/\hat{\gamma}$. The use of this limit yields the following terms of the expansion of the right-hand side in powers of ε. Defining

$$-\varepsilon \rho_\beta \mathcal{P} \frac{\mathcal{L}_2^{\dagger} \tilde{\rho}_\beta}{\tilde{\rho}_\beta} \equiv g = (g_0 + \varepsilon g_1) \, \rho_\beta ,$$

we have

$$g_0 = -\frac{1}{4} \left[U''(q) - \beta p^2 U''(q) \right] , \tag{5.36}$$

$$g_1 = -\frac{1}{12} \left[3\beta p U'(q) U''(q) - \beta p^3 U'''(q) + 6\beta p^4 - 28 p^2 + 10 \beta^{-1} \right] . \tag{5.37}$$

Furthermore, by equating powers of the reciprocal friction ε, we can solve a sequence of equations

$$\begin{aligned} \bar{\mathcal{L}}_{\mathrm{O}}^{\dagger} (\rho_\beta \hat{f}_{2,0}) &= g_0 \rho_\beta , \\ \mathcal{L}_{\mathrm{H}}^{\dagger} (\rho_\beta \hat{f}_{2,0}) + \bar{\mathcal{L}}_{\mathrm{O}}^{\dagger} (\rho_\beta \hat{f}_{2,1}) &= g_1 \rho_\beta , \\ \mathcal{L}_{\mathrm{H}}^{\dagger} (\rho_\beta \hat{f}_{2,1}) + \bar{\mathcal{L}}_{\mathrm{O}}^{\dagger} (\rho_\beta \hat{f}_{2,2}) &= 0 , \\ &\vdots \end{aligned}$$

to obtain the leading term $\hat{f}_{2,0}$, i.e.,

$$\hat{f}_{2,0} \equiv \hat{f}_{2,0}^{\mathrm{BADODAB}} = \frac{1}{8} \left(U''(q) - \beta p^2 U''(q) \right) . \tag{5.38}$$

Moreover, it can be easily shown that the marginal average of $\hat{f}_{2,0}^{\mathrm{BADODAB}}$ with respect to momentum is zero, i.e.,

$$\int_{\Omega_p} \hat{f}_{2,0}^{\mathrm{BADODAB}}(q,p) \rho_\beta \, \mathrm{d}\omega_p = 0 , \tag{5.39}$$

which leads to the average of configurational observables $\phi(q)$ with respect to the

invariant measure as

$$\langle \phi(q) \rangle_{\mathrm{BADODAB}} = \langle \phi(q) \rangle + h^2 \langle \phi(q) \hat{f}_{2,0}^{\mathrm{BADODAB}} \rangle + O(\varepsilon h^2 + h^4) \,.$$

Thus, for configurational observables the BADODAB method has fourth order convergence to the invariant measure in the large friction and thermal mass limits (i.e., $\varepsilon \to 0$),

$$\lim_{\varepsilon \to 0} \langle \phi(q) \rangle_{\mathrm{BADODAB}} = \langle \phi(q) \rangle + O(h^4) \,.$$

It should be emphasized here that only the BADODAB and BAODOAB methods appear to have the superconvergence property among a number of different splitting methods investigated in the Ad-Langevin/SGNHT system (5.16) with a clean gradient. The superconvergence property suggests the use of relatively large σ_{A} and $\mu \propto \sigma_{\mathrm{A}}^2$ in the BADODAB (SGNHT-S) method in order to enhance sampling accuracy. In fact, we expect that larger values of μ than this bound will not diminish the sampling accuracy, but the effect of large values of μ is to reduce the responsiveness of the thermostat device.

5.4 Numerical Experiments

In this section, we conduct a variety of numerical experiments to compare the performance of the different schemes presented in this chapter.

5.4.1 Molecular Systems

Before we compare various methods in machine learning applications (i.e., with a noisy gradient), we first demonstrate the order of convergence of various splitting methods with a clean gradient.

A popular model of an N-body system with pair interactions based on a spring with rest length (i.e., pendulum) (5.41) was used, a standard if simplified model of molecular dynamics. The total potential energy of the system is defined as

$$U(\mathbf{q}) = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \varphi(r_{ij}) \,, \tag{5.40}$$

where $r_{ij} = \|\mathbf{q}_i - \mathbf{q}_j\|$ denotes the distance between two particles i and j, and $\varphi(r_{ij})$ represents the pair potential energy

$$\varphi(r_{ij}) = \begin{cases} \dfrac{k}{2} \left(r_{ij} - r_{\mathrm{c}}\right)^2 \,, & r_{ij} < r_{\mathrm{c}} \,; \\ 0 \,, & r_{ij} \geq r_{\mathrm{c}} \,, \end{cases} \tag{5.41}$$

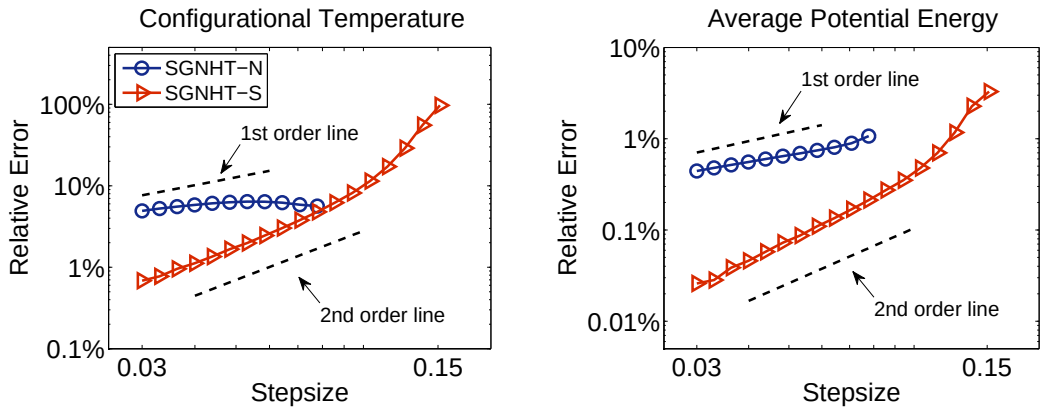


Figure 5.1: Log-log plot of the relative error in computed configurational temperature (left) and average potential energy (right) against stepsize by using two Ad-Langevin/SGNHT methods (with a clean gradient). The system ($\sigma_{\mathrm{A}} = 3$) was simulated for 5000 reduced time units, but only the last 80% of the data were collected to calculate the quantity to make sure the system was well equilibrated. Ten different runs were averaged to further reduce the sampling errors. The stepsizes tested began at $h = 0.03$ and were increased incrementally by 10% until both methods showed significant relative error (SGNHT-N became unstable at around $h = 0.08$).

where k and r_{c} represent the spring constant and the cutoff radius, respectively.

A system consisting of $N = 500$ identical particles (i.e., unit mass) was simulated in a cubic box with periodic boundary conditions [7]. The positions of the particles were initialized on a cubic grid with equidistant grid spacing, while the initial momenta were i.i.d. random variables with mean zero and variance $k_{\mathrm{B}}T$, which was set to be unity. The thermal mass μ was chosen to be 10 unless otherwise stated. Particle density $\rho_{\mathrm{d}} = 4$ was used with spring constant $k = 25$ and cutoff radius $r_{\mathrm{c}} = 1$.

We first compare the two SGNHT methods on controlling two configurational quantities: configurational temperature and average potential energy. As shown in Figure 5.1, with the help of the dashed order lines, we can see that SGNHT-N and SGNHT-S show first and second order convergence, respectively, as expected. It is clear that SGNHT-S has not only at least one order of magnitude improvement in accuracy in both observables, but also much greater robustness over the SGNHT-N method, which becomes completely unstable at around $h = 0.08$. The results on the configurational temperature and average potential energy are rather similar; therefore in what follows we present only average potential energy results.

We also investigate the effect of changing the value of σ_{A} in the SGNHT-S/BADODAB scheme proposed in this chapter. As can be seen from Figure 5.2, the SGNHT-S method displays second order convergence to the invariant measure when σ_{A} is relatively small, while a fourth order convergence is observed in the high friction limit ($\sigma_{\mathrm{A}} = 9$), as anticipated from the analysis of the previous section. It should be emphasized here that the superconvergence property was

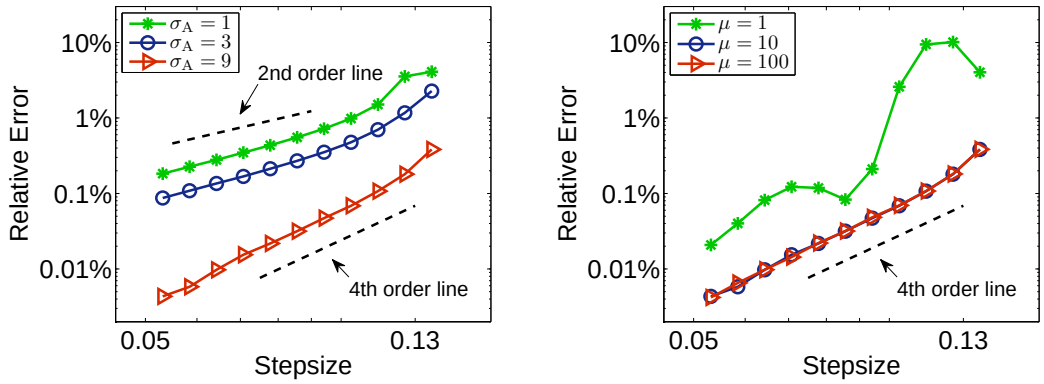


Figure 5.2: Log-log plot of the relative error in computed average potential energy against stepsize by using the SGNHT-S/BADODAB method with (left) different values of σ_A ($\mu = 10$) and (right) different values of μ ($\sigma_A = 9$). The format of the plots is the same as in Figure 5.1 except 50 different runs were used to reduce the sampling errors in high accuracy regime

.

observed only in the BADODAB and BAODOAB methods, which both reduce to the BAOAB method [99, 102] in Langevin dynamics.

Figure 5.2 also compares the effect of varying the value of the thermal mass μ when σ_A is fixed. It can be seen that the BADODAB method displays a clear fourth order convergence when μ is relatively large, while when μ is small, not only is the smooth discretization error dependence on stepsize lost, but significantly larger relative error is also observed. This reinforces the choice of a relatively large value of μ. It is worth pointing out that $\mu = 10$ works as well as $\mu = 100$; therefore $\mu = 10$ is used throughout this section since a relatively smaller μ corresponds to a tighter interaction between the thermostat and the system, and thus it can fluctuate more rapidly to accommodate changes in the noise and adapt more easily.

We also explore in Figure 5.3 the performance of various splitting methods of the Ad-Langevin/SGNHT system (5.16) with fixed values of σ_A and μ. All the methods clearly show second order convergence, with ABDODBA and BADODAB methods achieving one order of magnitude improvement in accuracy compared to the other methods. This again illustrates the importance of optimal design of numerical methods. The ABDODBA method seems to be slightly better that the BADODAB method in the regime of $\sigma_A = 3$; however, as demonstrated in Figure 5.2, the BADODAB method achieves a dramatic improvement in accuracy when σ_A is relatively large (e.g., $\sigma_A = 9$), while other schemes remain the same except for the BAODOAB method.

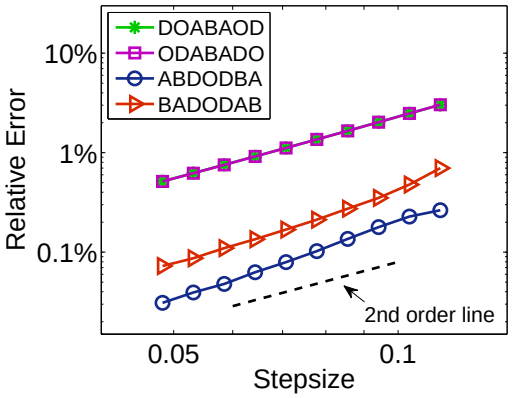


Figure 5.3: Log-log plot of the relative error in computed average potential energy against stepsize by using various splitting methods of the Ad-Langevin/SGNHT system ($\sigma_{\mathrm{A}} = 3$). The format of the plot is the same as in Figure 5.1.

5.4.2 Bayesian Inference

In this section we compare methods in a classical Bayesian inference model in one dimension, i.e., to estimate the mean of a normal distribution with known variance [44]. More precisely, given N i.i.d. samples from a normal distribution, $x_i \sim \mathcal{N}(\check{\mu}, \hat{\sigma}^2)$, where it should be noted that $\check{\mu}$ is the true mean, when we draw samples with known $\hat{\sigma}^2$ and a uniform prior distribution ranging from $-N/2$ to $N/2$, we are able to calculate the posterior distribution of the mean in a closed form

$$\hat{\mu} \sim \mathcal{N}\left(\hat{x}, \frac{\hat{\sigma}^2}{N}\right) , \tag{5.42}$$

where $\hat{x} = \sum_{i=1}^{N} x_i/N$. In the context of stochastic gradient approximation, we have

$$\begin{aligned}
\pi(\hat{\mu}|\mathbf{X}) \propto \pi(\mathbf{X}|\hat{\mu})\pi(\hat{\mu}) &\approx \left(\prod_{i=1}^{\tilde{n}} \pi(\mathbf{x}_{r_i}|\hat{\mu})\right)^{\frac{N}{\tilde{n}}} \pi(\hat{\mu}) \\
&= \left(\frac{1}{\sqrt{2\pi}\hat{\sigma}}\right)^{N} \left[\prod_{i=1}^{\tilde{n}} \exp\left(-\frac{(x_i-\hat{\mu})^2}{2\hat{\sigma}^2}\right)\right]^{\frac{N}{\tilde{n}}} \frac{1}{N} \\
&= \left(\frac{1}{\sqrt{2\pi}\hat{\sigma}}\right)^{N} \exp\left(-\frac{N}{\tilde{n}}\sum_{i=1}^{\tilde{n}} \frac{(x_i-\hat{\mu})^2}{2\hat{\sigma}^2}\right) \frac{1}{N} \\
&\propto \exp\left(-\frac{N}{\tilde{n}}\sum_{i=1}^{\tilde{n}} \frac{(x_i-\hat{\mu})^2}{2\hat{\sigma}^2}\right) \\
&= \exp\left[-\frac{1}{2\hat{\sigma}^2}\frac{N}{\tilde{n}}\left(\sum_{i=1}^{\tilde{n}}(x_i-\bar{x})^2 + \tilde{n}(\bar{x}-\hat{\mu})^2\right)\right] \\
&\propto \exp\left(-\frac{N}{2\hat{\sigma}^2}(\bar{x}-\hat{\mu})^2\right) ,
\end{aligned} \tag{5.43}$$

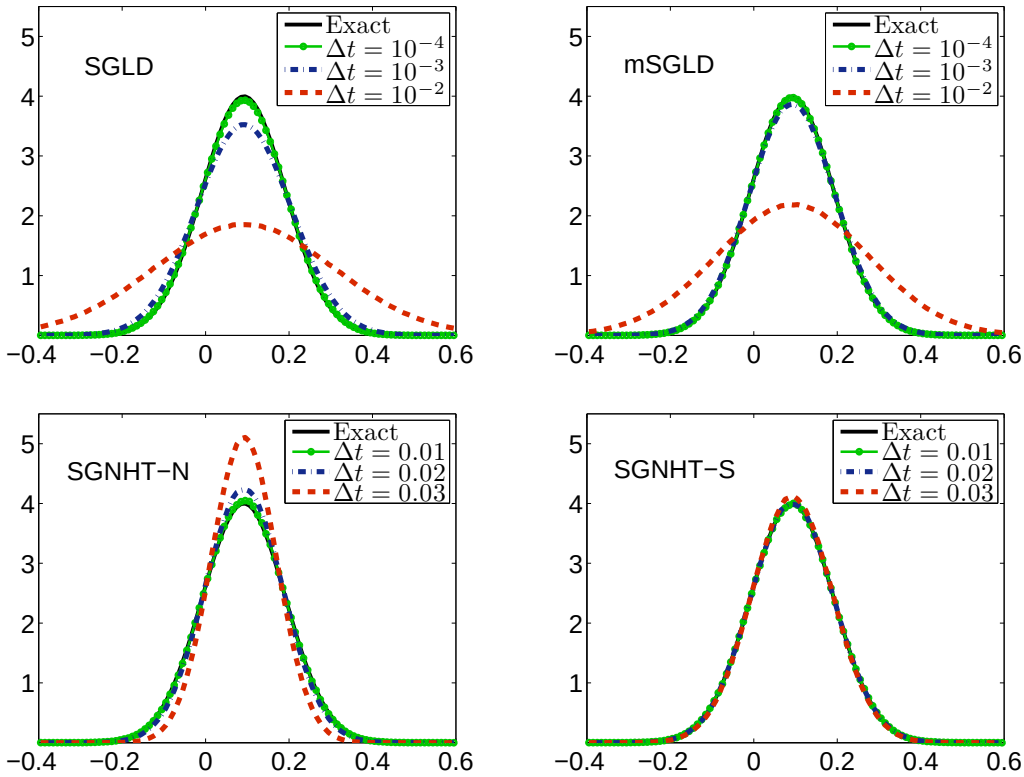


Figure 5.4: Comparisons of the distribution in a one-dimensional Bayesian inference problem by using SGLD (top left), mSGLD (top right), SGNHT-N (bottom left), and SGNHT-S (bottom right) with different stepsizes indicated by different colors. The solid black line is the exact solution. Note the difference in the legends between rows.

where $\bar{x} = \sum_{i=1}^{\tilde{n}} x_i/\tilde{n}$. It clearly recovers the true distribution (5.42) when $\tilde{n} = N$. Taking the logarithm and differentiating the posterior distribution obtained at the end of (5.43) with respect to $\hat{\mu}$ gives the noisy force

$$\tilde{F}(\hat{\mu}) = \frac{N}{\hat{\sigma}^2}\left(\hat{\mu} - \frac{1}{\tilde{n}}\sum_{i=1}^{\tilde{n}} x_i\right) . \tag{5.44}$$

In this simple case, the noise of the stochastic gradient is independent of $\hat{\mu}$ and is a constant given $\tilde{n}$. Moreover, we are able to obtain its mean and variance with respect to the stochastic gradient [78, 173]:

$$\begin{aligned} \mathbb{E}\tilde{F}(\hat{\mu}) &= F(\hat{\mu}) = \frac{N}{\hat{\sigma}^2}\left(\hat{\mu} - \frac{1}{N}\sum_{i=1}^{N} x_i\right) , \\ \mathrm{Var}\tilde{F}(\hat{\mu}) &= \frac{1}{\hat{\sigma}^4}\frac{N(N-1)}{\tilde{n}}\mathrm{Var}\mathbf{X} , \end{aligned} \tag{5.45}$$

where $\mathrm{Var}\mathbf{X}$ is the variance of the dataset. Thus, it is straightforward to verify that the noise is normally distributed according to the central limit theorem.

In our numerical experiments, σ_{A} was chosen as 1 due to the fact that large σ_{A} results in stability issues here. We generated $N = 100$ samples from $\mathcal{N}(0, 1)$ and randomly selected a subset of size $\tilde{n} = 10$ at each timestep to compute the noisy force (5.44). We plot the distributions of the posterior mean of the dataset obtained by using four different methods with different stepsizes in Figure 5.4. Clearly, two SGNHT methods completely outperformed the SGLD and mSGLD methods. The latter only demonstrate good approximation of the true distribution with order of magnitude smaller stepsize compared to the former. But it should be noted that mSGLD here is slightly better than SGLD in maintaining the true distribution: the distribution of mSGLD with $h = 0.001$ is visibly much closer to the target compared to that of SGLD with the same stepsize.

Note that stepsizes for SGNHT (second order dynamics) and SGLD (first order dynamics) based methods are not directly comparable—as mentioned in [99] the stepsize of a first order dynamics method like Euler–Maruyama when viewed as the limiting discretization of a Langevin integrator corresponds to $h^2/2$, where h is the stepsize of the Langevin method. However, in our experiments we are uninterested in the time-dynamics of the system and care only about the invariant measure. Therefore the important relationship is the error in thermodynamic averages in comparison with the number of timesteps (work), which quantifies the efficiency of a given method. The stepsize is just an arbitrary parameter which allows for refinement of the statistical calculation.

Between the two SGNHT methods, SGNHT-S (the new scheme being proposed here) is obviously superior to SGNHT-N: the latter starts to show significant deviation from the true distribution at $h = 0.02$, while the distribution of the former still looks well matched to the true one at $h = 0.03$. Our observations are confirmed by Figure 5.5, where the mean absolute error (MAE) of the distribution of the two SGNHT methods is plotted. The MAE, which can be thought of as a relative error in distribution, is defined as

$$\mathrm{MAE} = \frac{1}{\bar{N}} \sum_{i=1}^{\bar{N}} \left| \omega_i - \hat{\omega}_i \right|, \tag{5.46}$$

where $\bar{N}$ denotes the number of intervals, which was chosen as 100. ω_i and $\hat{\omega}_i$ represent the observed frequency in bin i and the exact expected frequency, respectively [99]. As can be seen, the stability threshold of SGNHT-N was around $h = 0.03$, beyond which the system became unstable, as highlighted in the figure (in which case the system blew up, resulting in a 100% MAE). Once again, SGNHT-S not only shows an order of magnitude better accuracy but also has a much greater robustness than SGNHT-N. In particular, for defined accuracy, the SGNHT-S method is able to use double the stepsize compared to SGNHT-N, which means a remarkable 50% improvement in overall numerical efficiency as defined in Section 3.7.4.

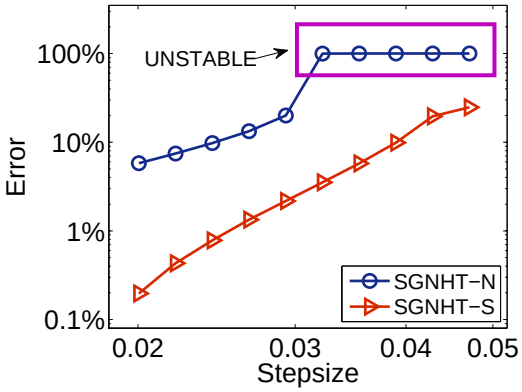


Figure 5.5: Log-log plot of the mean absolute error (MAE) in the distribution of the Bayesian inference model against stepsize. The box indicates that the system was unstable with corresponding stepsizes for the SGNHT-N method.

5.4.3 Bayesian Logistic Regression

Following [173], we also investigate the performance of different methods for a more complicated Bayesian logistic regression model. The data $y_i \in \{-1, 1\}$ were modelled by

$$\pi(y_i|\mathbf{x}_i, \boldsymbol{\beta}) = f(y_i \boldsymbol{\beta}^T \mathbf{x}_i)\,,$$

where $f(z) = 1/(1+\exp(-z)) \in [0,1]$ is the logistic function and $\mathbf{x}_i \in \mathbb{R}^d$ are rows of a fixed dataset. Our goal is to estimate the posterior mean of parameter vector $\boldsymbol{\beta} \in \mathbb{R}^d$. For simplicity, a multivariate Gaussian prior $\mathcal{N}(\mathbf{0}, \mathbf{I})$ was used on $\boldsymbol{\beta}$. Therefore, by using Bayes' theorem, we obtain the following posterior distribution:

$$\pi(\boldsymbol{\beta}) \propto \exp\left(-\frac{1}{2}\|\boldsymbol{\beta}\|^2\right) \prod_{i=1}^{N} f(y_i \boldsymbol{\beta}^T \mathbf{x}_i)\,. \tag{5.47}$$

Following the same procedure in the Bayesian inference example (Section 5.4.2), we can calculate the noisy force and then plug it into different thermostats for sampling.

In our numerical experiments, we considered the $d = 3$ case with $N = 1000$ data points. We chose the dataset to be

$$\mathbf{X} = \begin{pmatrix} x_{1,1} & x_{1,2} & 1 \\ x_{2,1} & x_{2,2} & 1 \\ \vdots & \vdots & \vdots \\ x_{1000,1} & x_{1000,2} & 1 \end{pmatrix}, \tag{5.48}$$

where $x_{i,j}$ were sampled from a standard normal distribution $\mathcal{N}(0,1)$ for $i = 1, \ldots, 1000$ and $j = 1, 2$. A subset of size $\tilde{n} = 100$ was randomly chosen at each timestep to compute the noisy force.

The performance of estimating the posterior mean value of parameter vector $\boldsymbol{\beta}$

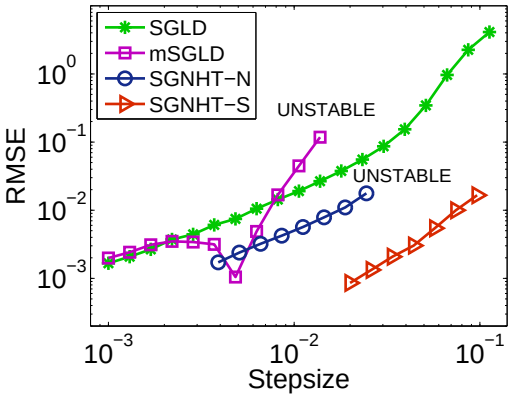


Figure 5.6: Comparisons of the root mean square error (RMSE) of the posterior mean in the Bayesian logistic regression model by using various methods against stepsize. The system was simulated for 1000 reduced time units with 100,000 different runs. The stepsizes tested began at $h = 0.001$ and were increased incrementally by 30% until all methods either displayed significant error or became unstable (mSGLD and SGNHT-N).

by various methods ($\sigma_{\mathrm{A}} = 6$) was tested and plotted in Figure 5.6. Again, SGLD and mSGLD, displaying considerably larger root mean square error (RMSE) with a fixed stepsize, were outperformed by the two SGNHT methods. In this case, the SGLD and mSGLD methods demonstrated similar control in numerical accuracy, but the latter displayed much worse stability than that of the former and became unstable just above $h = 0.01$. As reported in the original paper [173], the performance of the mSGLD method depends strongly on the size of the subset—for a larger subset, which requires higher computational cost, the bias of mSGLD can be smaller than that of SGLD.

Of the two SGNHT methods, the SGNHT-S method again shows not only at least an order of magnitude improvement on accuracy but also much better robustness than the other: SGNHT-N became unstable just above $h = 0.02$. Remarkably, the SGNHT-S method at $h = 0.1$ still achieves better accuracy than the SGLD method at $h = 0.01$. In other words, the method we propose here gives more than a 90% improvement in overall numerical efficiency compared to one of the most popular methods in the literature. For fixed accuracy, the SGNHT-S method can use almost four times the stepsize of the SGNHT-N method (i.e., an improvement of about 75% in overall numerical efficiency).

5.5 Covariance-Controlled Adaptive Langevin (CCAdL) Thermostat

Ding et al. [44] claimed that it is reasonable to assume the covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is constant when the size of the dataset, N, is large, in which case the variance

of the posterior of $\boldsymbol{\theta}$ is small. The magnitude of the posterior variance does not actually relate to the constancy of the $\boldsymbol{\Sigma}$, however, in general, $\boldsymbol{\Sigma}$ is not constant. Simply assuming the nonconstancy of the $\boldsymbol{\Sigma}$ can have a significant impact on the performance of the method (most notably the stability measured by the largest usable stepsize). Therefore, it is essential to have an approach that can handle parameter-dependent noise. In the following section, we propose a covariance-controlled thermostat that can effectively dissipate parameter-dependent noise while maintaining the target stationary distribution.

5.5.1 Formulation of CCAdL

As mentioned in the previous section, the SGNHT method (5.16) can only dissipate noise with a constant covariance matrix. When the covariance matrix becomes parameter-dependent, in general, a parameter-dependent covariance matrix does not imply the required "thermal equilibrium", i.e., the system cannot be expected to converge to the desired invariant distribution (5.17), typically resulting in poor estimation of functions of parameters of interest. In fact, in that case it is not clear whether or not there exists an invariant distribution at all.

In order to construct a stochastic-dynamical system that preserves the canonical distribution, we suggest adding a suitable damping (viscous) term to effectively dissipate the parameter-dependent gradient noise. To this end, we propose the following covariance-controlled adaptive Langevin (CCAdL) thermostat:

$$\begin{aligned}
\mathrm{d}\boldsymbol{\theta} &= \mathbf{M}^{-1}\mathbf{p}\mathrm{d}t\,,\\
\mathrm{d}\mathbf{p} &= -\nabla U(\boldsymbol{\theta})\mathrm{d}t + \sqrt{h\boldsymbol{\Sigma}(\boldsymbol{\theta})}\mathbf{M}^{1/2}\mathrm{d}\mathbf{W} - (h/2)\beta\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{p}\mathrm{d}t - \xi\mathbf{p}\mathrm{d}t + \sqrt{2A\beta^{-1}}\mathbf{M}^{1/2}\mathrm{d}\mathbf{W}_{\mathrm{A}}\,,\\
\mathrm{d}\xi &= \mu^{-1}\left[\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p} - N_{\mathrm{d}}k_{\mathrm{B}}T\right]\mathrm{d}t\,,
\end{aligned} \tag{5.49}$$

where A, which can be termed the "effective friction", is a positive parameter and proportional to the variance of the noise.

Proposition 2. *The CCAdL thermostat* (5.49) *preserves the modified Gibbs (stationary) distribution:*

$$\hat{\rho}_\beta(\boldsymbol{\theta},\mathbf{p},\xi) = \frac{1}{Z}\exp\left(-\beta H(\boldsymbol{\theta},\mathbf{p})\right)\exp\left(-\frac{\beta\mu}{2}(\xi - A)^2\right)\,. \tag{5.50}$$

Proof. The Fokker–Planck equation corresponding to (5.49) is

$$\begin{aligned}
\rho_t = \mathcal{L}^\dagger\rho := &-\mathbf{M}^{-1}\mathbf{p}\cdot\nabla_{\boldsymbol{\theta}}\rho + \nabla U(\boldsymbol{\theta})\cdot\nabla_{\mathbf{p}}\rho + (h/2)\nabla_{\mathbf{p}}\cdot(\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{M}\nabla_{\mathbf{p}}\rho)\\
&+ (h/2)\beta\nabla_{\mathbf{p}}\cdot(\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{p}\rho) + \xi\nabla_{\mathbf{p}}\cdot(\mathbf{p}\rho) + A\beta^{-1}\nabla_{\mathbf{p}}\cdot(\mathbf{M}\nabla_{\mathbf{p}}\rho)\\
&- \mu^{-1}\left[\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p} - N_{\mathrm{d}}k_{\mathrm{B}}T\right]\nabla_\xi\rho\,.
\end{aligned}$$

Just insert $\hat{\rho}_\beta$ (5.50) into the Fokker–Planck operator $\mathcal{L}^\dagger$ to see that it vanishes. □

The incorporation of the parameter-dependent covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ in (5.49) is intended to offset the covariance matrix coming from the gradient approximation. However, in practice, one does not know $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ a priori. Thus instead one must estimate $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ during the simulation, a task which will be addressed below. This procedure is related to the method used in the SGHMC method proposed by Chen et al. [32], which uses dynamics of the following form:

$$\begin{aligned}
&\mathrm{d}\boldsymbol{\theta} = \mathbf{M}^{-1}\mathbf{p}\mathrm{d}t\,,\\
&\mathrm{d}\mathbf{p} = -\nabla U(\boldsymbol{\theta})\mathrm{d}t + \sqrt{h\boldsymbol{\Sigma}(\boldsymbol{\theta})}\mathbf{M}^{1/2}\mathrm{d}\mathbf{W} - A\mathbf{p}\mathrm{d}t + \sqrt{2\beta^{-1}\left(A\mathbf{I} - \beta h\boldsymbol{\Sigma}(\boldsymbol{\theta})/2\right)}\mathbf{M}^{1/2}\mathrm{d}\mathbf{W}_{\mathrm{A}}\,.
\end{aligned}$$

It can be shown that the SGHMC method preserves the Gibbs canonical distribution:

$$\rho_\beta(\boldsymbol{\theta}, \mathbf{p}) = Z^{-1}\exp\left(-\beta H(\boldsymbol{\theta}, \mathbf{p})\right)\,. \tag{5.51}$$

Although both CCAdL (5.49) and SGHMC (5.5.1) preserve their respective invariant distributions, let us note several advantages of the former over the latter in practice:

(i) CCAdL and SGHMC both require estimation of the covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ during simulation, which can be costly in high dimension. In numerical experiments, we have found that simply using the diagonal of the covariance matrix, at significantly reduced computational cost, works quite well in CCAdL. By contrast, it is difficult to find a suitable value of the parameter A in SGHMC since one has to make sure the matrix $A\mathbf{I} - \beta h\boldsymbol{\Sigma}(\boldsymbol{\theta})/2$ is positive semidefinite. One may attempt to use a large value of the "effective friction" A and/or a small stepsize h. However, too-large a friction would essentially reduce SGHMC to SGLD, which is not desirable, as pointed out in [32], while extremely small stepsize would significantly impact the computational efficiency.

(ii) Estimation of the covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ unavoidably introduces additional noise in both CCAdL and SGHMC. Nonetheless, CCAdL can still effectively control the system temperature (i.e., maintaining the correct distribution of the momenta) due to the use of the stabilizing Nosé–Hoover control, while in SGHMC, poor estimation of the covariance matrix may lead to significant deviations of the system temperature (as well as the distribution of the momenta), resulting in poor sampling of the parameters of interest.

Covariance Estimation of Noisy Gradients

Under the assumption that the noise of the stochastic gradient follows a normal distribution, we apply a similar method to that of [4] to estimate the covariance

matrix associated with the noisy gradient. If we let $g(\boldsymbol{\theta}; \mathbf{x}) = \nabla_{\boldsymbol{\theta}} \log \pi(\mathbf{x}|\boldsymbol{\theta})$ and assume that the size of subset $\tilde{n}$ is large enough for the central limit theorem to hold, we have

$$\frac{1}{\tilde{n}} \sum_{i=1}^{\tilde{n}} g(\boldsymbol{\theta}_t; \mathbf{x}_{r_i}) \sim \mathcal{N}\left(\mathbb{E}_{\mathbf{x}}[g(\boldsymbol{\theta}_t; \mathbf{x})], \frac{1}{\tilde{n}} \mathbf{I}_t \right) , \tag{5.52}$$

where $\mathbf{I}_t = \mathrm{Cov}[g(\boldsymbol{\theta}_t; \mathbf{x})]$ is the covariance of the gradient at $\boldsymbol{\theta}_t$. Given the noisy (stochastic) gradient based on the current subset

$$\nabla \tilde{U}(\boldsymbol{\theta}_t) = -\frac{N}{\tilde{n}} \sum_{i=1}^{\tilde{n}} g(\boldsymbol{\theta}_t; \mathbf{x}_{r_i}) - \nabla \log \pi(\boldsymbol{\theta}_t) \tag{5.53}$$

and the clean (full) gradient

$$\nabla U(\boldsymbol{\theta}_t) = -\sum_{i=1}^{N} g(\boldsymbol{\theta}_t; \mathbf{x}_i) - \nabla \log \pi(\boldsymbol{\theta}_t) , \tag{5.54}$$

we have

$$\mathbb{E}_{\mathbf{x}}[\nabla \tilde{U}(\boldsymbol{\theta}_t)] = \mathbb{E}_{\mathbf{x}}[\nabla U(\boldsymbol{\theta}_t)] , \tag{5.55}$$

and thus

$$\nabla \tilde{U}(\boldsymbol{\theta}_t) = \nabla U(\boldsymbol{\theta}_t) + \mathcal{N}\left(\mathbf{0}, \frac{N^2}{\tilde{n}} \mathbf{I}_t \right) , \tag{5.56}$$

i.e., $\boldsymbol{\Sigma}(\boldsymbol{\theta}_t) = N^2 \mathbf{I}_t / \tilde{n}$. Assuming $\boldsymbol{\theta}_t$ does not change dramatically over time, we use the moving average update to estimate $\mathbf{I}_t$:

$$\hat{\mathbf{I}}_t = (1 - \kappa_t) \hat{\mathbf{I}}_{t-1} + \kappa_t \mathbf{V}(\boldsymbol{\theta}_t) , \tag{5.57}$$

where $\kappa_t = 1/t$ and

$$\mathbf{V}(\boldsymbol{\theta}_t) = \frac{1}{\tilde{n} - 1} \sum_{i=1}^{\tilde{n}} \left(g(\boldsymbol{\theta}_t; \mathbf{x}_{r_i}) - \bar{g}(\boldsymbol{\theta}_t) \right) \left(g(\boldsymbol{\theta}_t; \mathbf{x}_{r_i}) - \bar{g}(\boldsymbol{\theta}_t) \right)^T \tag{5.58}$$

is the empirical covariance of the gradient. $\bar{g}(\boldsymbol{\theta}_t)$ represents the mean gradient of the log likelihood computed from a subset. As proved in [4], this estimator has a convergence order of $\mathcal{O}(1/N)$.

As already mentioned, estimating the full covariance matrix is computationally infeasible in high dimension. However, we have found that employing a diagonal approximation of the covariance matrix (i.e., estimating the variance only along each dimension of the noisy gradient) works quite well in practice, as demonstrated in Section 5.5.2.

The procedure of the CCAdL method is summarized in Algorithm 1, where we simply used $\mathbf{M} = \mathbf{I}$, $\beta = 1$, and $\mu = N_{\mathrm{d}}$ in order to be consistent with the original implementation of SGNHT [44].

Algorithm 1 Covariance-Controlled Adaptive Langevin (CCAdL) Thermostat

Input: h, A, $\{\kappa_t\}_{t=1}^{\hat{T}}$.
Initialize $\boldsymbol{\theta}_0$, $\mathbf{p}_0$, $\mathbf{I}_0$, and $\xi_0 = A$.
for $t = 1, 2, \ldots, \hat{T}$ **do**
  $\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \mathbf{p}_{t-1}h$;
  Estimate $\hat{\mathbf{I}}_t$ using Equation (5.57);
  $\mathbf{p}_t = \mathbf{p}_{t-1} - \nabla\tilde{U}(\boldsymbol{\theta}_t)h - \frac{h}{2}\frac{N^2}{\tilde{n}}\hat{\mathbf{I}}_t\mathbf{p}_{t-1}h - \xi_{t-1}\mathbf{p}_{t-1}h + \sqrt{2Ah}\mathcal{N}(0, \mathbf{I})$;
  $\xi_t = \xi_{t-1} + \left(\mathbf{p}_t^T\mathbf{p}_t/N_{\mathrm{d}} - 1\right)h$;
end for

Note that this is a simple, first order (in terms of the stepsize) algorithm. We have introduced in Section 5.3 higher order of accuracy schemes which can improve accuracy, but our interest here is in the direct comparison of the underlying machinery of SGHMC, SGNHT, and CCAdL, so we avoid further modifications and enhancements related to timestepping at this stage.

In the following section, we compare the newly established CCAdL method with SGHMC and SGNHT on various machine learning tasks to demonstrate the benefits of CCAdL in Bayesian sampling with a noisy gradient.

5.5.2 Numerical Experiments

Bayesian Inference for Gaussian Distribution

We first compare the performance of the newly established CCAdL method with SGHMC and SGNHT for a simple task using synthetic data, i.e., Bayesian inference of both the mean and variance of a one-dimensional normal distribution. We apply the same experimental setting as in [44]. We generated $N = 100$ samples from a standard normal distribution $\mathcal{N}(0, 1)$. We used the likelihood function of $\mathcal{N}(\mathbf{x}_i|\mu, \gamma^{-1})$ and assigned a normal-gamma distribution as their prior distribution, i.e., $\mu, \gamma \sim \mathcal{N}(\mu|0, \gamma)\mathrm{Gamma}(\gamma|1, 1)$. Then the corresponding posterior distribution is another normal-gamma distribution, i.e.,

$$(\mu, \gamma)|\mathbf{X} \sim \mathcal{N}(\mu|\mu_N, (\kappa_N\gamma)^{-1})\mathrm{Gamma}(\gamma|\alpha_N, \beta_N)\,, \tag{5.59}$$

with

$$\mu_N = \frac{N\bar{\mathbf{x}}}{N+1}\,, \quad \kappa_N = 1+N\,, \quad \alpha_N = 1+\frac{N}{2}\,, \quad \beta_N = 1+\sum_{i=1}^{N}\frac{(\mathbf{x}_i - \bar{\mathbf{x}})^2}{2}+\frac{N\bar{\mathbf{x}}^2}{2(1+N)}\,,$$

where $\bar{\mathbf{x}} = \sum_{i=1}^{N}\mathbf{x}_i/N$. A random subset of size $\tilde{n} = 10$ was selected at each timestep to approximate the full gradient, resulting in the following stochastic

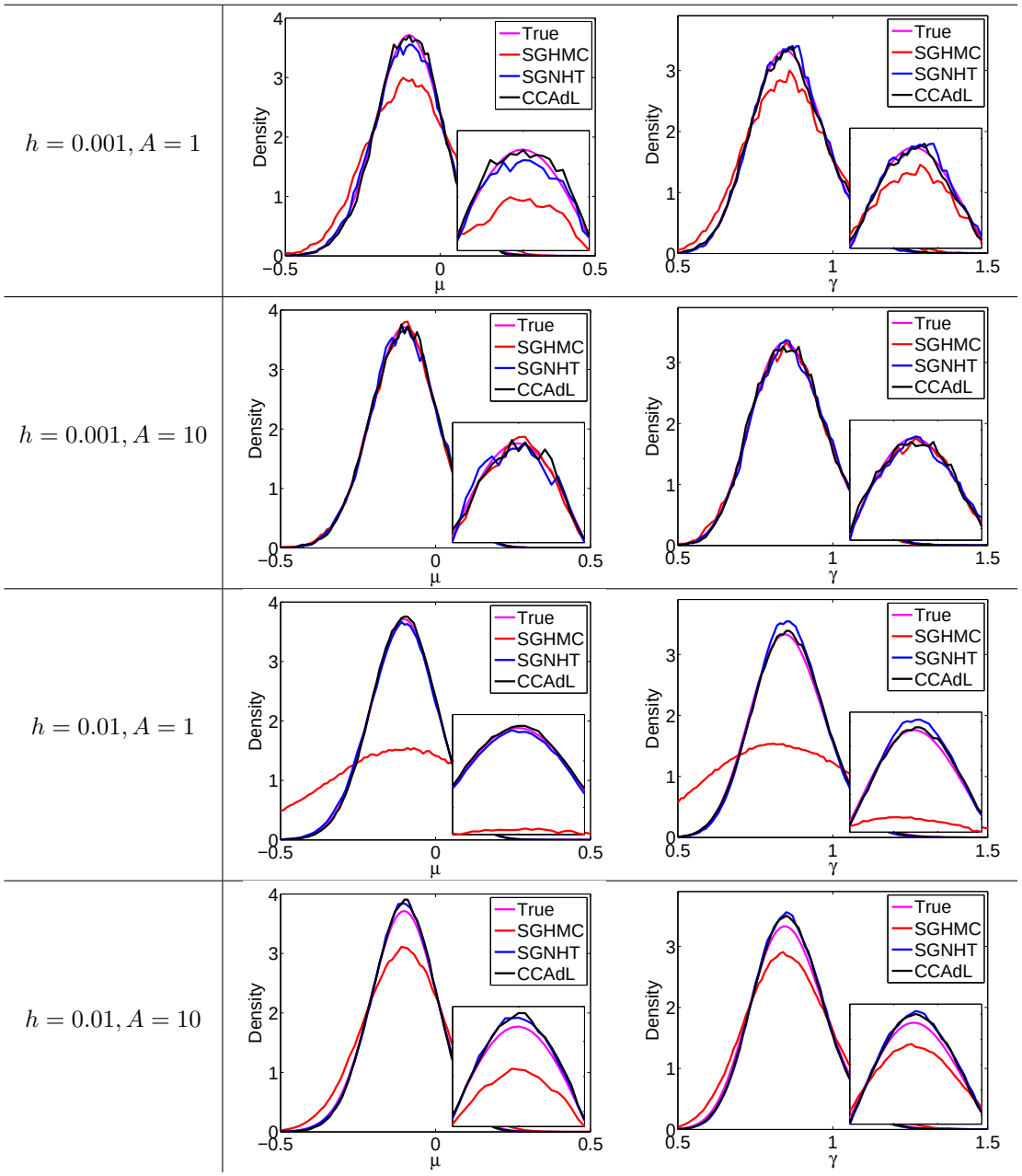


Figure 5.7: Comparisons of marginal distribution (density) of μ (top row) and γ (bottom row) with various values of h and A indicated in each column. The peak region is highlighted in the inset.

gradients [124]:

$$\nabla_\mu \tilde{U} = (N+1)\mu\gamma - \frac{\gamma N}{n}\sum_{i=1}^{n}\mathbf{x}_{r_i}\,, \qquad \nabla_\gamma \tilde{U} = 1 - \frac{N+1}{2\gamma} + \frac{\mu^2}{2} + \frac{N}{2n}\sum_{i=1}^{n}(\mathbf{x}_{r_i} - \mu)^2\,.$$

It can be seen that the variance of the stochastic gradient noise is no longer constant and actually depends on the size of the subset, $\tilde{n}$, and the values of μ and γ in each iteration. This directly violates the constant noise variance assumption of SGNHT [44], while CCAdL adjusts to the varying noise variance.

The marginal distributions of μ and γ obtained from various methods with different combinations of h and A were compared and plotted in Figure 5.7, with Table 5.1 consisting of the corresponding root mean square error (RMSE) of the distribution and autocorrelation time [125] from 10^6 samples. In most of the cases, both SGNHT and CCAdL easily outperform the SGHMC method possibly due to the presence of the Nosé–Hoover device, with SGHMC only showing superiority with small value of h and large value of A, neither of which is desirable in practice as discussed in Section 5.5.1. Between SGNHT and the newly proposed CCAdL method, the latter achieves better performance in each of the cases investigated, highlighting the importance of the covariance control with parameter-dependent noise.

Method	$h = 0.001, A = 1$	$h = 0.001, A = 10$	$h = 0.01, A = 1$	$h = 0.01, A = 10$
SGHMC	$(0.0148, 236.12)$	$(\mathbf{0.0029}, \mathbf{333.04})$	$(0.0531, 29.78)$	$(0.0132, 39.33)$
SGNHT	$(0.0037, 238.32)$	$(0.0035, 406.71)$	$(0.0044, 26.71)$	$(0.0043, 55.00)$
CCAdL	$(\mathbf{0.0034}, \mathbf{238.06})$	$(0.0031, 402.45)$	$(\mathbf{0.0021}, \mathbf{26.71})$	$(\mathbf{0.0035}, \mathbf{54.43})$

Table 5.1: Comparisons of (RMSE, Autocorrelation time) of (μ, γ) of various methods for Bayesian inference of the mean and variance of a Gaussian distribution.

Large-Scale Bayesian Logistic Regression

We then consider a Bayesian logistic regression model trained on the benchmark MNIST dataset for binary classification of digits 7 and 9 using $12,214$ training data points, with a test set of size 2037. A 100-dimensional random projection of the original features was used. We used the likelihood function of

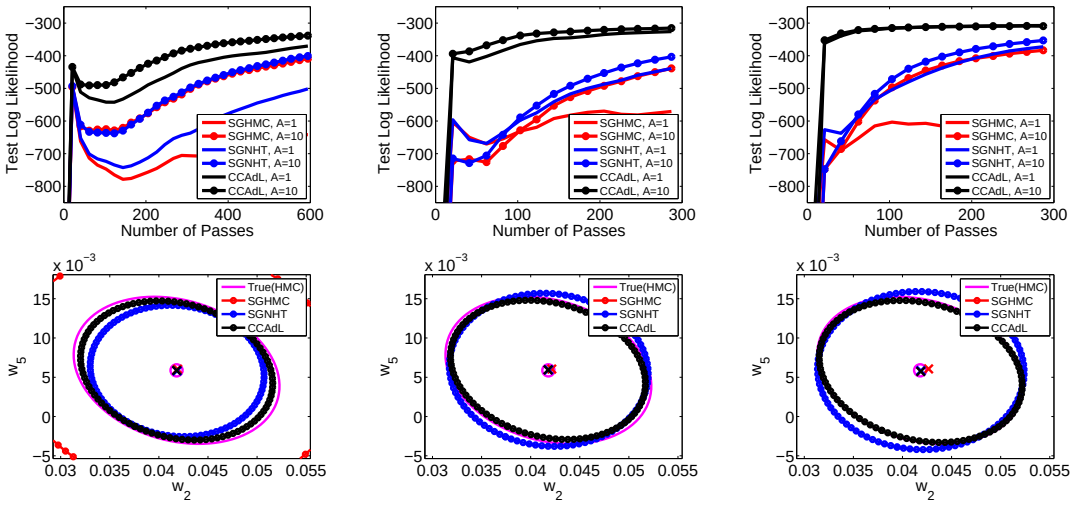
$$\pi\left(\{\mathbf{x}_i, y_i\}_{i=1}^N | \mathbf{w}\right) \propto \prod_{i=1}^N 1/\left(1 + \exp(-y_i \mathbf{w}^T \mathbf{x}_i)\right) \tag{5.60}$$

and the prior distribution of

$$\pi(\mathbf{w}) \propto \exp(-\mathbf{w}^T \mathbf{w}/2)\,. \tag{5.61}$$

A subset of size $\tilde{n} = 500$ was used at each timestep. Since the dimensionality of this problem is not that high, a full covariance estimation was used for CCAdL.

We investigate in Figure 5.8 (top row) the convergence speed of each method through measuring test log likelihood using the posterior mean against the number of passes over the entire dataset. CCAdL displays significant improvements over SGHMC and SGNHT with different values of h and A: (1) CCAdL converges much faster than the other two, which also indicates its faster mixing speed and shorter burn-in period; (2) CCAdL shows robustness in different values of the effective friction A, with SGHMC and SGNHT relying on a relative large value of A (especially for the SGHMC method), which is intended to dominate the gradient noise.



(a) $h = 0.2 \times 10^{-4}$ (b) $h = 0.5 \times 10^{-4}$ (c) $h = 1 \times 10^{-4}$

Figure 5.8: Comparisons of Bayesian logistic regression of various methods on the MNIST dataset of digits 7 and 9 with various values of h and A: (top row) test log likelihood using the posterior mean against the number of passes over the entire dataset; (bottom row) two-dimensional marginal posterior distribution in (randomly selected) dimensions 2 and 5 with $A = 10$ fixed, based on 10^6 samples from each method after the burn-in period (i.e., we start to collect samples when the test log likelihood stabilizes). Magenta circle is the true (reference) posterior mean obtained from standard HMC, and crosses represent the sample means computed from various methods. Ellipses represent isoprobability contours covering 95% probability mass. Note that the contour of SGHMC is way beyond the scale of the plot especially in the large stepsize regime, in which case we do not include it here.

To compare the sample quality obtained from each method, Figure 5.8 (bottom row) plots the two-dimensional marginal posterior distribution in randomly selected dimensions of 2 and 5 based on 10^6 samples from each method after the burn-in period (i.e., we start to collect samples when the test log likelihood stabilizes). The true (reference) distribution was obtained by a sufficiently long run of standard HMC. We implemented 10 runs of standard HMC and found there was no variation between these runs, which guarantees its qualification as the true (reference) distribution. Again, CCAdL shows much better performance than SGHMC and SGNHT. Note that the contour of SGHMC does not even fit in the region of the plot, and in fact it shows significant deviation even in the estimation of the mean.

Chapter 6

Conclusions

We have reviewed a number of numerical methods that are widely used in DPD simulations and have also proposed a new stochastic momentum-conserving thermostat, the pairwise Nosé–Hoover–Langevin (PNHL) thermostat. Two favorable splitting methods of the PNHL thermostat were introduced and compared to existing methods in the computation of various physical quantities (both static and dynamical).

We have observed that, for the PNHL thermostat proposed here, the PNHL-N method based on a nonsymmetric arrangement of the terms of a splitting, gives an enormous stability benefit. The PNHL-N method needs to calculate the force twice in each integration step, which is computationally costly in the model setting used in this thesis; nevertheless when the computational overhead is costed carefully the PNHL-N method outperforms the alternatives. To measure the practical performance of numerical methods in DPD simulations quantitatively, we have defined the "numerical efficiency", based on which we have reported substantial improvements of both methods of the newly proposed thermostat, with the symmetric PNHL-S method 46% more efficient than the commonly used DPD-VV method and the nonsymmetric PNHL-N method incredibly 87% better than the benchmark method in DPD simulations. It should be noted that, if the force calculation is not that expensive in other model settings, the gain of using the PNHL-N method could be further exploited.

Based on the numerical experiments of the velocity and transverse momentum autocorrelation functions which characterize the translational and rotational diffusions of the system respectively, DPD and PNHL give rather similar dynamical properties in practice. Although PNHL as formulated is not based on a hydrodynamic interaction model [72, 89], we have seen that it is an effective replacement for DPD in the low-friction regime. Moreover, we point out that the projection along interacting particle pairs in PNHL could be replaced by alternatives to achieve further control of transport properties [86, 109, 110]. The method is also potentially useful more broadly in molecular simulation applications, whenever momentum conservation is at issue.

The only difference between PNH and PNHL is the additional Langevin thermostat acting on the dynamical friction, however they perform very differently. This is probably because the cutoffs used in PNH and the simple potentials there provided insufficient internal mechanisms for the system to achieve an ergodic sampling of the canonical distribution. In molecular dynamics with steep pair potentials (e.g., Lennard-Jones), the ergodic properties develop more naturally and the "L" in PNHL can be redundant in some instances. It is also worth mentioning that both PNH and PNHL are not able to vary the Schmidt number since the average of the dynamical friction tends to zero, whereas the Schmidt number can be tuned in some of the other schemes (see also recent work based on the Mori–Zwanzig formulation [109, 110]).

We have also investigated the order of convergence of the long-time averages to the invariant measure for a couple of methods described in this thesis. By extending the framework recently introduced in Langevin dynamics, we can infer (and verify using numerics) the second order convergence for those nonsymmetric GLA-like methods (DPD-S1, LA, and Peters) as well as the nonsymmetric PNHL-N method. However, rigorous investigation on other nonsymmetric methods (DPD-VV, NHLA, and PNH) that surprisingly obtained second order convergence remains to be established. Overall, we claim here that the PNHL thermostat indeed can be used (and may be preferred in some typical cases) as an alternative to low-friction DPD simulations with substantially improved the computational efficiency and no degradation of convergence rate.

We have proposed a novel momentum-conserving pairwise adaptive Langevin (PAdL) thermostat, which is applicable in DPD and related systems described in this thesis. Unlike the PNHL thermostat, the PAdL thermostat is able to fully capture the dynamics as in DPD. Furthermore, we have observed substantial improvements in numerical efficiency over alternative schemes, i.e., an at least 130% improvement over a wide range of friction coefficients and shear rates in equilibrium and nonequilibrium simulations, respectively.

We have reviewed a variety of methods in stochastic gradient systems with applications in machine learning. We have provided a theoretical discussion on the foundation (underlying dynamics) of those stochastic gradient systems, which has been lacking in the literature. We have also proposed a new symmetric splitting (at least second order) method in SGNHT (SGNHT-S/BADODAB), which substantially improves the accuracy and robustness compared to a nonsymmetric splitting (first order) method (SGNHT-N) proposed recently in the literature. Furthermore, we have demonstrated that under certain conditions the SGNHT-S/BADODAB method can inherit the superconvergence property recently discovered in integrators for Langevin dynamics, i.e., fourth order convergence to the invariant measure for configurational averages.

By conducting various numerical experiments, we have demonstrated that the two SGNHT methods outperform the popular SGLD method and its variant mSGLD. In particular, the SGNHT-S method can use up to ten times the stepsize

of SGLD, which implies a remarkable more than 90% improvement in overall numerical efficiency. Between the two SGNHT methods, the SGNHT-S method can use almost four times the stepsize of SGNHT-N for defined accuracy (i.e., about a 75% improvement in overall numerical efficiency).

It should be noted that in certain cases, it may be desirable to employ a Metropolis–Hastings procedure in order to remove the discretization bias [148]. However, we emphasize that the correction is not without computational cost, particularly as the dimension is increased [15, 88, 146, 147], and the results of [99, 100, 102] and of the current article demonstrate that high accuracy with respect to the invariant distribution is often achievable using traditional numerical integration techniques, thus in many cases entirely eliminating the necessity of Metropolis–Hastings corrections (see more discussions in [102]). Moreover, we mention that the methods of this article can in principle be combined with Metropolis–Hastings algorithms if it is necessary to completely eliminate the discretization bias.

We have proposed a novel covariance-controlled adaptive Langevin (CCAdL) formulation that can effectively dissipate parameter-dependent noise while maintaining a desired invariant distribution. CCAdL combines ideas of SGHMC and SGNHT from the literature, but achieves significant improvements over each of these methods in practice. The additional error introduced by covariance estimation is expected to be small in a relative sense, i.e., substantially smaller than the error arising from the noisy gradient. Our findings have been verified in large-scale machine learning applications. In particular, we have consistently observed that SGHMC relies on a small stepsize h and a large friction A, which significantly reduces its usefulness in practice as discussed. The techniques presented could be of use in more general settings of large-scale Bayesian sampling and optimization, which we leave for future work.

A naive nonsymmetric splitting method has been applied for CCAdL for fair comparison in this thesis. However, we point out that it is desirable to investigate the optimal design of splitting methods in noisy gradient systems with parameter-dependent noise. We leave further exploration of this direction with data science applications for future work.

Finally, we highlight the key contributions of this thesis as follows:

- we have introduced a new formulation that combines a pairwise Nosé–Hoover thermostat with the Langevin-type stochastic thermostatting in auxiliary variables, i.e., the pairwise Nosé–Hoover–Langevin (PNHL) thermostat, which is shown to be a functional alternative to dissipative particle dynamics (DPD) with good control of thermodynamic and dynamic properties in the low-friction regime;

- we have systematically compared various formulations and numerical algorithms, and have demonstrated a numerical scheme (the PNHL-N method)

which offers greater than an 80% efficiency enhancement compared to the most popular algorithm for DPD simulation;

- we have proposed a novel momentum-conserving thermostat, i.e., the pairwise adaptive Langevin (PAdL) thermostat, that fully captures the dynamics of DPD;
- we have demonstrated that the newly proposed PAdL thermostat is able to achieve a more than 130% improvements in numerical efficiency over a wide range of friction coefficients and shear rates in equilibrium and nonequilibrium simulations, respectively;
- we have reviewed and compared various formulations for noisy gradient simulations where the force is polluted by random perturbations of zero mean and unknown variance;
- we have provided a theoretical discussion of the foundations of those methods based on discussions of the corresponding underlying dynamics;
- we have proposed a new numerical method for the adaptive Langevin/stochastic gradient Nosé–Hoover thermostat, i.e., the BADODAB/SGNHT-S method, that achieves an impressive improvement in numerical efficiency over the most popular stochastic gradient methods reported in the literature;
- we have demonstrated that the newly established method inherits a superconvergence property (fourth order convergence to its invariant measure for configurational quantities) recently demonstrated in the setting of Langevin dynamics;
- we have proposed a covariance-controlled adaptive Langevin (CCAdL) thermostat that can effectively dissipate parameter-dependent noise while maintaining a desired target distribution;
- we have demonstrated in large-scale machine learning applications that the newly proposed method achieves a substantial speedup over popular alternative schemes.

Bibliography

[1] A. Abdulle, G. Vilmart, and K. C. Zygalakis. High order numerical approximation of the invariant measure of ergodic SDEs. *SIAM Journal on Numerical Analysis*, 52(4):1600–1622, 2014.

[2] A. Abdulle, G. Vilmart, and K. C. Zygalakis. Long time accuracy of Lie–Trotter splitting methods for Langevin dynamics. *SIAM Journal on Numerical Analysis*, 53(1):1–16, 2015.

[3] Y. Afshar, F. Schmid, A. Pishevar, and S. Worley. Exploiting seeding of random number generators for efficient domain decomposition parallelization of dissipative particle dynamics. *Computer Physics Communications*, 184(4):1119–1128, 2013.

[4] S. Ahn, A. Korattikara, and M. Welling. Bayesian posterior sampling via stochastic gradient Fisher scoring. In *Proceedings of the 29th International Conference on Machine Learning*, pages 1591–1598, 2012.

[5] M. P. Allen. Configurational temperature in membrane simulations using dissipative particle dynamics. *The Journal of Physical Chemistry B*, 110 (8):3823–3830, 2006.

[6] M. P. Allen and F. Schmid. A thermostat for molecular dynamics of complex fluids. *Molecular Simulation*, 33(1-2):21–26, 2007.

[7] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, 1989.

[8] H. C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72(4):2384, 1980.

[9] G. Arya, E. J. Maginn, and H.-C. Chang. Efficient viscosity estimation from molecular dynamics simulation via momentum impulse relaxation. *The Journal of Chemical Physics*, 113(6):2079, 2000.

[10] R. B. Ash. *Basic Probability Theory*. Dover Publications, 2008.

[11] W. T. Ashurst and W. G. Hoover. Dense-fluid shear viscosity via nonequilibrium molecular dynamics. *Physical Review A*, 11(2):658–678, 1975.

[12] J. A. Backer, C. P. Lowe, H. C. J. Hoefsloot, and P. D. Iedema. Poiseuille flow to measure the viscosity of particle model fluids. *The Journal of Chemical Physics*, 122:154503, 2005.

[13] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81:3684, 1984.

[14] B. J. Berne and G. D. Harp. On the calculation of time correlation functions. *Advances in Chemical Physics*, pages 63–227, 1970.

[15] A. Beskos, N. Pillai, G. Roberts, J.-M. Sanz-Serna, and A. Stuart. Optimal tuning of the hybrid Monte Carlo algorithm. *Bernoulli*, 19(5A):1501–1534, 2013.

[16] G. Besold, I. Vattulainen, M. Karttunen, and J. M. Polson. Towards better integrators for dissipative particle dynamics simulations. *Physical Review E*, 62(6):R7611, 2000.

[17] E. S. Boek, P. V. Coveney, and H. N. W. Lekkerkerker. Computer simulation of rheological phenomena in dense colloidal suspensions with dissipative particle dynamics. *Journal of Physics: Condensed Matter*, 8(47): 9509–9512, 1996.

[18] E. S. Boek, P. V. Coveney, H. N. W. Lekkerkerker, and P. van der Schoot. Simulating the rheology of dense colloidal suspensions using dissipative particle dynamics. *Physical Review E*, 55(3):3124, 1997.

[19] M. Born and T. von Karman. Über schwingungen von raumgittern. *Physikalische Zeitschrift*, 13:297–309, 1912.

[20] F. A. Bornemann, P. Nettesheim, and C. Schütte. Quantum-classical molecular dynamics as an approximation to full quantum dynamics. *The Journal of Chemical Physics*, 105(3):1074–1083, 1996.

[21] A. Boromand, S. Jamali, and J. M. Maia. Viscosity measurement techniques in dissipative particle dynamics. *Computer Physics Communications*, 196: 149–160, 2015.

[22] N. Bou-Rabee and H. Owhadi. Long-run accuracy of variational integrators in the stochastic context. *SIAM Journal on Numerical Analysis*, 48(1):278–297, 2010.

[23] C. Braga and K. P. Travis. A configurational temperature Nosé–Hoover thermostat. *The Journal of Chemical Physics*, 123:134101, 2005.

[24] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng. *Handbook of Markov Chain Monte Carlo*. CRC Press, 2011.

[25] A. Brünger, C. L. Brooks III, and M. Karplus. Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chemical Physics Letters*, 105(5):495–500, 1984.

[26] G. Bussi and M. Parrinello. Accurate sampling using Langevin dynamics. *Physical Review E*, 75(5):056707, 2007.

[27] B. Butler, G. Ayton, O. G. Jepps, and D. J. Evans. Configurational temperature: Verification of Monte Carlo simulations. *The Journal of Chemical Physics*, 109(16):6519, 1998.

[28] E. Cancès, F. Legoll, and G. Stoltz. Theoretical and numerical comparison of some sampling methods for molecular dynamics. *ESAIM: Mathematical Modelling and Numerical Analysis*, 41(2):351–389, 2007.

[29] S. Chakravorty. A homotopic Galerkin approach to the solution of the Fokker–Planck–Kolmogorov equation. In *Proceedings of the 2006 American Control Conference*. IEEE, 2006.

[30] A. Chatterjee. Modification to Lees–Edwards periodic boundary condition for dissipative particle dynamics simulation with high dissipation rates. *Molecular Simulation*, 33(15):1233–1236, 2007.

[31] A. Chaudhri and J. R. Lukes. Velocity and stress autocorrelation decay in isothermal dissipative particle dynamics. *Physical Review E*, 81(2):026707, 2010.

[32] T. Chen, E. B. Fox, and C. Guestrin. Stochastic gradient Hamiltonian Monte Carlo. In *Proceedings of the 31st International Conference on Machine Learning*, pages 1683–1691, 2014.

[33] G. Ciccotti, G. Jacucci, and I. R. McDonald. "Thought-experiments" by molecular dynamics. *Journal of Statistical Physics*, 21(1):1–22, 1979.

[34] G. Ciccotti and A. Tenenbaum. Canonical ensemble and nonequilibrium states by molecular dynamics. *Journal of Statistical Physics*, 23(6):767–772, 1980.

[35] C. J. Cotter and S. Reich. An extended dissipative particle dynamics model. *Europhysics Letters*, 64(6):723–729, 2003.

[36] M. Dashti and A. M. Stuart. The Bayesian approach to inverse problems. *arXiv preprint arXiv:1302.6989*, 2013.

[37] G. De Fabritiis, M. Serrano, P. Español, and P. Coveney. Efficient numerical integrators for stochastic models. *Physica A: Statistical Mechanics and its Applications*, 361(2):429–440, 2006.

[38] A. Debussche and E. Faou. Weak backward error analysis for SDEs. *SIAM Journal on Numerical Analysis*, 50(3):1735–1752, 2012.

[39] J. Delhommelle and D. J. Evans. Configurational temperature profile in confined fluids. I. Atomic fluid. *The Journal of Chemical Physics*, 114(14): 6229, 2001.

[40] J. Delhommelle and D. J. Evans. Configurational temperature profile in confined fluids. II. Molecular fluids. *The Journal of Chemical Physics*, 114 (14):6236, 2001.

[41] J. Delhommelle and D. J. Evans. Configurational temperature thermostat for fluids undergoing shear flow: application to liquid chlorine. *Molecular Physics*, 99(21):1825–1829, 2001.

[42] J. Delhommelle and D. J. Evans. Correspondence between configurational temperature and molecular kinetic temperature thermostats. *The Journal of Chemical Physics*, 117(13):6016, 2002.

[43] W. den Otter and J. Clarke. A new algorithm for dissipative particle dynamics. *Europhysics Letters*, 53(4):426, 2001.

[44] N. Ding, Y. Fang, R. Babbush, C. Chen, R. D. Skeel, and H. Neven. Bayesian sampling using stochastic gradient thermostats. In *Advances in Neural Information Processing Systems 27*, pages 3203–3211, 2014.

[45] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222, 1987.

[46] S. Dubinkina, J. Frank, and B. Leimkuhler. Simplified modelling of a thermal bath, with application to a fluid vortex system. *Multiscale Modeling & Simulation*, 8(5):1882–1901, 2010.

[47] A. Einstein. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 322(8):549–560, 1905.

[48] D. L. Ermak and H. Buckholz. Numerical integration of the Langevin equation: Monte Carlo simulation. *Journal of Computational Physics*, 35 (2):169–182, 1980.

[49] D. L. Ermak and J. McCammon. Brownian dynamics with hydrodynamic interactions. *The Journal of Chemical Physics*, 69:1352, 1978.

[50] P. Español and P. Warren. Statistical mechanics of dissipative particle dynamics. *Europhysics Letters*, 30(4):191, 1995.

[51] D. J. Evans and G. Morriss. *Statistical Mechanics of Nonequilibrium Liquids*. Cambridge University Press, 2008.

[52] D. J. Evans and G. P. Morriss. Shear thickening and turbulence in simple fluids. *Physical Review Letters*, 56(20):2172, 1986.

[53] X. Fan, N. Phan-Thien, S. Chen, X. Wu, and T. Y. Ng. Simulating flow of DNA suspension using dissipative particle dynamics. *Physics of Fluids*, 18 (6):063102, 2006.

[54] D. A. Fedosov and G. E. Karniadakis. Triple-decker: Interfacing atomistic-mesoscopic-continuum flow regimes. *Journal of Computational Physics*, 228 (4):1157–1171, 2009.

[55] D. A. Fedosov, G. E. Karniadakis, and B. Caswell. Steady shear rheometry of dissipative particle dynamics models of polymer fluids in reverse Poiseuille flow. *The Journal of Chemical Physics*, 132(14):144103, 2010.

[56] D. A. Fedosov, H. Noguchi, and G. Gompper. Multiscale modeling of blood flow: from single cells to blood rheology. *Biomechanics and Modeling in Mechanobiology*, 13(2):239–258, 2013.

[57] D. A. Fedosov, W. Pan, B. Caswell, G. Gompper, and G. E. Karniadakis. Predicting human blood viscosity in silico. *Proceedings of the National Academy of Sciences*, 108(29):11772–11777, 2011.

[58] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications, Second Edition*. Academic Press, 2001.

[59] U. Frisch, B. Hasslacher, and Y. Pomeau. Lattice-gas automata for the Navier-Stokes equation. *Physical Review Letters*, 56(14):1505–1508, 1986.

[60] J. Gibson, K. Chen, and S. Chynoweth. The equilibrium of a velocity-Verlet type algorithm for DPD with finite time steps. *International Journal of Modern Physics C*, 10(01):241–261, 1999.

[61] D. Givon, R. Kupferman, and A. Stuart. Extracting macroscopic dynamics: model problems and algorithms. *Nonlinearity*, 17(6):R55–R127, 2004.

[62] N. Goga, A. Rzepiela, A. de Vries, S. Marrink, and H. Berendsen. Efficient algorithms for Langevin and DPD dynamics. *Journal of Chemical Theory and Computation*, 8(10):3637–3649, 2012.

[63] M. S. Green. Markoff random processes and the statistical mechanics of time-dependent phenomena. II. Irreversible processes in fluids. *The Journal of Chemical Physics*, 22(3):398–413, 1954.

[64] R. D. Groot. Applications of Dissipative Particle Dynamics. In *Novel Methods in Soft Matter Simulations*, pages 5–38. Springer, 2004.

[65] R. D. Groot and P. B. Warren. Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation. *The Journal of Chemical Physics*, 107:4423, 1997.

[66] E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*. Springer, 2006.

[67] Y. Han and D. G. Grier. Configurational temperature of charge-stabilized colloidal monolayers. *Physical Review Letters*, 92(14):148301, 2004.

[68] Y. Han and D. G. Grier. Configurational temperatures and interactions in charge-stabilized colloid. *The Journal of Chemical Physics*, 122(6):064907, 2005.

[69] J.-P. Hansen and I. McDonald. *Theory of Simple Liquids, Third Edition*. Academic Press, 2006.

[70] M. W. Heemels, C. P. Lowe, and A. F. Bakker. The wavelength dependence of the high-frequency shear viscosity in a colloidal suspension of hard spheres. *Progress in Colloid & Polymer Science*, 110:150–155, 1998.

[71] B. Hess. Determining the shear viscosity of model liquids from molecular dynamics simulations. *The Journal of Chemical Physics*, 116(1):209–217, 2002.

[72] C. Hijón, P. Español, E. Vanden-Eijnden, and R. Delgado-Buscalioni. Mori–Zwanzig formalism as a practical computational tool. *Faraday Discussions*, 144:301–322, 2010.

[73] M. D. Homan and A. Gelman. The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *The Journal of Machine Learning Research*, 15(1):1593–1623, 2014.

[74] P. Hoogerbrugge and J. Koelman. Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics. *Europhysics Letters*, 19(3): 155, 1992.

[75] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31(3):1695, 1985.

[76] W. G. Hoover. *Computational Statistical Mechanics, Studies in Modern Thermodynamics*. Elsevier Science, 1991.

[77] A. M. Horowitz. A generalized guided Monte Carlo algorithm. *Physics Letters B*, 268(2):247–252, 1991.

[78] D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, 1952.

[79] P. H. Hünenberger. Thermostat algorithms for molecular dynamics simulations. *Advances in Polymer Science*, 173:105–149, 2005.

[80] A. F. Jakobsen, G. Besold, and O. G. Mouritsen. Multiple time step update schemes for dissipative particle dynamics. *The Journal of Chemical Physics*, 124:094104, 2006.

[81] A. F. Jakobsen, O. G. Mouritsen, and G. Besold. Artifacts in dynamical simulations of coarse-grained model lipid bilayers. *The Journal of Chemical Physics*, 122(20):204901, 2005.

[82] S. Jamali, A. Boromand, S. Khani, and J. Maia. Gaussian-inspired auxiliary non-equilibrium thermostat (GIANT) for dissipative particle dynamics simulations. *Computer Physics Communications*, 197:27–34, 2015.

[83] O. G. Jepps, G. Ayton, and D. J. Evans. Microscopic expressions for the thermodynamic temperature. *Physical Review E*, 62(4):4757–4763, 2000.

[84] A. Jones and B. Leimkuhler. Adaptive stochastic methods for sampling driven molecular systems. *The Journal of Chemical Physics*, 135(8):084125, 2011.

[85] R. Joubaud and G. Stoltz. Nonequilibrium shear viscosity computations with Langevin dynamics. *Multiscale Modeling & Simulation*, 10(1):191–216, 2012.

[86] C. Junghans, M. Praprotnik, and K. Kremer. Transport properties controlled by a thermostat: An extended dissipative particle dynamics thermostat. *Soft Matter*, 4(1):156–161, 2008.

[87] E. E. Keaveny, I. V. Pivkin, M. Maxey, and G. E. Karniadakis. A comparative study between dissipative particle dynamics and molecular dynamics for simple- and complex-geometry flows. *The Journal of Chemical Physics*, 123:104107, 2005.

[88] A. D. Kennedy and B. Pendleton. Acceptances and autocorrelations in hybrid Monte Carlo. *Nuclear Physics B - Proceedings Supplements*, 20: 118–121, 1991.

[89] T. Kinjo and S.-a. Hyodo. Equation of motion for coarse-grained simulation based on microscopic description. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, 75(5 Pt 1):051109–051109, 2007.

[90] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, 1992.

[91] E. A. Koopman and C. P. Lowe. Advantages of a Lowe–Andersen thermostat in molecular dynamics simulations. *The Journal of Chemical Physics*, 124(20):204103, 2006.

[92] R. Kubo. Statistical-mechanical theory of irreversible processes. I. General theory and simple applications to magnetic and conduction problems. *Journal of the Physical Society of Japan*, 12(6):570–586, 1957.

[93] J. P. Larentzos, J. K. Brennan, J. D. Moore, M. Lísal, and W. D. Mattson. Parallel implementation of isothermal and isoenergetic Dissipative Particle Dynamics using Shardlow-like splitting algorithms. *Computer Physics Communications*, 185(7):1987–1998, 2014.

[94] A. Lees and S. Edwards. The computer study of transport processes under extreme conditions. *Journal of Physics C: Solid State Physics*, 5(15):1921, 1972.

[95] F. Legoll, M. Luskin, and R. Moeckel. Non-ergodicity of the Nosé–Hoover thermostatted harmonic oscillator. *Archive for Rational Mechanics and Analysis*, 184(3):449–463, 2006.

[96] F. Legoll, M. Luskin, and R. Moeckel. Non-ergodicity of Nosé–Hoover dynamics. *Nonlinearity*, 22(7):1673–1694, 2009.

[97] B. Leimkuhler. Generalized Bulgac–Kusnezov methods for sampling of the Gibbs–Boltzmann measure. *Physical Review E*, 81(2):026703, 2010.

[98] B. Leimkuhler, D. T. Margul, and M. E. Tuckerman. Stochastic, resonance-free multiple time-step algorithm for molecular dynamics with very large time steps. *Molecular Physics*, 111(22-23):3579–3594, 2013.

[99] B. Leimkuhler and C. Matthews. Rational construction of stochastic numerical methods for molecular sampling. *Applied Mathematics Research eXpress*, 2013(1):34–56, 2013.

[100] B. Leimkuhler and C. Matthews. Robust and efficient configurational molecular sampling via Langevin dynamics. *The Journal of Chemical Physics*, 138:174102, 2013.

[101] B. Leimkuhler and C. Matthews. *Molecular Dynamics: With Deterministic and Stochastic Numerical Methods*. Springer, 2015.

[102] B. Leimkuhler, C. Matthews, and G. Stoltz. The computation of averages from equilibrium and nonequilibrium Langevin molecular dynamics. *IMA Journal of Numerical Analysis*, 36(1):13–79, 2016.

[103] B. Leimkuhler, C. Matthews, and M. V. Tretyakov. On the long-time integration of stochastic gradient systems. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 470(2170):20140120, 2014.

[104] B. Leimkuhler, E. Noorizadeh, and F. Theil. A gentle stochastic thermostat for molecular dynamics. *Journal of Statistical Physics*, 135(2):261–277, 2009.

[105] B. Leimkuhler and S. Reich. *Simulating Hamiltonian Dynamics*. Cambridge University Press, 2005.

[106] B. Leimkuhler and X. Shang. On the numerical treatment of dissipative particle dynamics and related systems. *Journal of Computational Physics*, 280:72–95, 2015.

[107] B. Leimkuhler and X. Shang. Adaptive thermostats for noisy gradient systems. *SIAM Journal on Scientific Computing*, 38(2):A712–A736, 2016.

[108] X. Li, P. M. Vlahovska, and G. E. Karniadakis. Continuum- and particle-based modeling of shapes and dynamics of red blood cells in health and disease. *Soft Matter*, 9(1):28–37, 2013.

[109] Z. Li, X. Bian, B. Caswell, and G. E. Karniadakis. Construction of dissipative particle dynamics models for complex fluids via the Mori–Zwanzig formulation. *Soft Matter*, 10(43):8659–8672, 2014.

[110] Z. Li, X. Bian, X. Li, and G. E. Karniadakis. Incorporation of memory effects in coarse-grained modeling via the Mori–Zwanzig formalism. *The Journal of Chemical Physics*, 143(24):243128, 2015.

[111] M. Lísal and J. K. Brennan. Alignment of lamellar diblock copolymer phases under shear: Insight from dissipative particle dynamics simulations. *Langmuir*, 23(9):4809–4818, 2007.

[112] M. Lísal, J. K. Brennan, and J. B. Avalos. Dissipative particle dynamics at isothermal, isobaric, isoenergetic, and isoenthalpic conditions using Shardlow-like splitting algorithms. *The Journal of Chemical Physics*, 135 (20):204105–204105, 2011.

[113] R. López-Esparza, M. A. Balderas Altamirano, E. Pérez, and A. Gama Goicochea. Importance of molecular interactions in colloidal dispersions. *Advances in Condensed Matter Physics*, 2015:1–8, 2015.

[114] C. Lowe. An alternative approach to dissipative particle dynamics. *Europhysics Letters*, 47(2):145, 1999.

[115] L. Lue, O. G. Jepps, J. Delhommelle, and D. J. Evans. Configurational thermostats for molecular systems. *Molecular Physics*, 100(14):2387–2395, 2002.

[116] J. C. Mattingly, A. M. Stuart, and D. J. Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic Processes and their Applications*, 101(2):185–232, 2002.

[117] S. Melchionna. Design of quasisymplectic propagators for Langevin dynamics. *The Journal of Chemical Physics*, 127:044108, 2007.

[118] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[119] G. N. Milstein and M. V. Tretyakov. *Stochastic Numerics for Mathematical Physics*. Springer, 2004.

[120] E. Moeendarbary, T. Ng, and M. Zangeneh. Dissipative particle dynamics: Introduction, methodology and complex fluid applications - a review. *International Journal of Applied Mechanics*, 1(04):737–763, 2009.

[121] L. Mones, A. Jones, A. W. Götz, T. Laino, R. C. Walker, B. Leimkuhler, G. Csányi, and N. Bernstein. The adaptive buffered force QM/MM method in the CP2K and AMBER software packages. *Journal of Computational Chemistry*, 36(9):633–648, 2015.

[122] A. Moshfegh and A. Jabbarzadeh. Modified Lees–Edwards boundary condition for dissipative particle dynamics: hydrodynamics and temperature at high shear rates. *Molecular Simulation*, 41(15):1264–1277, 2014.

[123] F. Müller-Plathe. Reversing the perturbation in nonequilibrium molecular dynamics: An easy way to calculate the shear viscosity of fluids. *Physical Review E*, 59(5):4894, 1999.

[124] K. P. Murphy. Conjugate Bayesian analysis of the Gaussian distribution. Technical report, 2007.

[125] R. M. Neal. Probabilistic inference using Markov chain Monte Carlo methods. Technical report, 1993.

[126] P. Nikunen, M. Karttunen, and I. Vattulainen. How would you integrate the equations of motion in dissipative particle dynamics simulations? *Computer Physics Communications*, 153(3):407–423, 2003.

[127] S. Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52(2):255–268, 1984.

[128] S. Nosé. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics*, 81(1):511, 1984.

[129] I. Pagonabarraga, M. Hagen, and D. Frenkel. Self-consistent dissipative particle dynamics algorithm. *Europhysics Letters*, 42(4):377, 1998.

[130] B. J. Palmer. Transverse-current autocorrelation-function calculations of the shear viscosity for molecular liquids. *Physical Review E*, 49(1):359, 1994.

[131] D. Pan, N. Phan-Thien, and B. C. Khoo. Dissipative particle dynamics simulation of droplet suspension in shear flow at low Capillary number. *Journal of Non-Newtonian Fluid Mechanics*, 212:63–72, 2014.

[132] W. Pan, B. Caswell, and G. E. Karniadakis. Rheology, microstructure and migration in Brownian colloidal suspensions. *Langmuir*, 26(1):133–142, 2010.

[133] C. Pastorino, T. Kreer, M. Müller, and K. Binder. Comparison of dissipative particle dynamics and Langevin thermostats for out-of-equilibrium simulations of polymeric systems. *Physical Review E*, 76(2):026706, 2007.

[134] S. Patterson and Y. W. Teh. Stochastic gradient Riemannian Langevin dynamics on the probability simplex. In *Advances in Neural Information Processing Systems 26*, pages 3102–3110, 2013.

[135] B. L. Peters, A. Ramírez-Hernández, D. Q. Pike, M. Müller, and J. J. de Pablo. Nonequilibrium simulations of lamellae forming block copolymers under steady shear: A comparison of dissipative particle dynamics and Brownian dynamics. *Macromolecules*, 45(19):8109–8116, 2012.

[136] E. Peters. Elimination of time step effects in DPD. *Europhysics Letters*, 66 (3):311, 2004.

[137] C. L. Phillips, J. A. Anderson, and S. C. Glotzer. Pseudo-random number generation for Brownian Dynamics and Dissipative Particle Dynamics simulations on GPU devices. *Journal of Computational Physics*, 230(19): 7191–7201, 2011.

[138] I. V. Pivkin, B. Caswell, and G. Karniadakis. Dissipative particle dynamics. *Reviews in Computational Chemistry*, 27:85–110, 2010.

[139] I. V. Pivkin and G. E. Karniadakis. A new method to impose no-slip boundary conditions in dissipative particle dynamics. *Journal of Computational Physics*, 207(1):114–128, 2005.

[140] M. Praprotnik, L. Delle Site, and K. Kremer. Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly. *The Journal of Chemical Physics*, 123(22):224106, 2005.

[141] M. Praprotnik, L. Delle Site, and K. Kremer. Multiscale simulation of soft matter: From scale bridging to adaptive resolution. *Annual Review of Physical Chemistry*, 59:545–571, 2008.

[142] M. Revenga, I. Zúñiga, and P. Español. Boundary conditions in dissipative particle dynamics. *Computer Physics Communications*, 121-122:309–311, 1999.

[143] H. Risken. *The Fokker–Planck Equation: Methods of Solution and Applications, Second Edition*. Springer, 1989.

[144] H. Robbins and S. Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, 22(2):400–407, 1951.

[145] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods, Second Edition*. Springer, 2004.

[146] G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120, 1997.

[147] G. O. Roberts and J. S. Rosenthal. Optimal scaling of discrete approximations to Langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1):255–268, 1998.

[148] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.

[149] H. H. Rugh. Dynamical approach to temperature. *Physical Review Letters*, 78(5):772, 1997.

[150] A. A. Samoletov, C. P. Dettmann, and M. A. J. Chaplain. Thermostats for "slow" configurational modes. *Journal of Statistical Physics*, 128(6): 1321–1336, 2007.

[151] A. A. Samoletov, C. P. Dettmann, and M. A. J. Chaplain. Notes on configurational thermostat schemes. *The Journal of Chemical Physics*, 132: 246101, 2010.

[152] M. A. Seaton, R. L. Anderson, S. Metz, and W. Smith. DL_MESO: highly scalable mesoscale simulations. *Molecular Simulation*, 39(10):796–821, 2013.

[153] M. Serrano, G. De Fabritiis, P. Español, and P. V. Coveney. A stochastic Trotter integration scheme for dissipative particle dynamics. *Mathematics and Computers in Simulation*, 72(2):190–194, 2006.

[154] X. Shang, Z. Zhu, B. Leimkuhler, and A. J. Storkey. Covariance-controlled adaptive Langevin thermostat for large-scale Bayesian sampling. In *Advances in Neural Information Processing Systems 28*, pages 37–45, 2015.

[155] T. Shardlow. Splitting for dissipative particle dynamics. *SIAM Journal on Scientific Computing*, 24(4):1267–1282, 2003.

[156] T. Shardlow and Y. Yan. Geometric ergodicity for dissipative particle dynamics. *Stochastics and Dynamics*, 6(01):123–154, 2006.

[157] R. D. Skeel and J. A. Izaguirre. An impulse integrator for Langevin dynamics. *Molecular Physics*, 100(24):3885–3891, 2002.

[158] T. Soddemann, B. Dünweg, and K. Kremer. Dissipative particle dynamics: A useful thermostat for equilibrium and nonequilibrium molecular dynamics simulations. *Physical Review E*, 68(4):046702, 2003.

[159] L. Stella, C. D. Lorenz, and L. Kantorovich. Generalized Langevin equation: An efficient approach to nonequilibrium molecular dynamics of open systems. *Physical Review B*, 89(13):134303, 2014.

[160] S. D. Stoyanov and R. D. Groot. From molecular dynamics to hydrodynamics: A novel Galilean invariant thermostat. *The Journal of Chemical Physics*, 122:114112, 2005.

[161] G. Strang. On the construction and comparison of difference schemes. *SIAM Journal on Numerical Analysis*, 5(3):506–517, 1968.

[162] V. Symeonidis, G. Karniadakis, and B. Caswell. A seamless approach to multiscale complex fluid simulation. *Computing in Science and Engineering*, 7(3):39–46, 2005.

[163] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Analysis and Applications*, 8(4):483–509, 1990.

[164] S. Tang, G. T. Evans, C. P. Mason, and M. P. Allen. Shear viscosity for fluids of hard ellipsoids: A kinetic theory and molecular dynamics study. *The Journal of Chemical Physics*, 102(9):3794–3811, 1995.

[165] Y. W. Teh, A. Thiéry, and S. Vollmer. Consistency and fluctuations for stochastic gradient Langevin dynamics. *arXiv preprint arXiv:1409.0578*, 2014.

[166] F. Thalmann and J. Farago. Trotter derivation of algorithms for Brownian and dissipative particle dynamics. *The Journal of Chemical Physics*, 127: 124109, 2007.

[167] K. P. Travis and C. Braga. Configurational temperature control for atomic and molecular systems. *The Journal of Chemical Physics*, 128:014111, 2008.

[168] H. F. Trotter. On the product of semi-groups of operators. *Proceedings of the American Mathematical Society*, 10(4):545–551, 1959.

[169] C. Trozzi and G. Ciccotti. Stationary nonequilibrium states by molecular dynamics. II. Newton's law. *Physical Review A*, 29(2):916–925, 1984.

[170] W. F. van Gunsteren and H. J. C. Berendsen. A leap-frog algorithm for stochastic dynamics. *Molecular Simulation*, 1(3):173–185, 1988.

[171] I. Vattulainen, M. Karttunen, G. Besold, and J. M. Polson. Integration schemes for dissipative particle dynamics simulations: From softly interacting systems towards hybrid models. *The Journal of Chemical Physics*, 116:3967, 2002.

[172] L. Verlet. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical Review*, 159(1):98, 1967.

[173] S. J. Vollmer, K. C. Zygalakis, and Y. W. Teh. (Non-) asymptotic properties of stochastic gradient Langevin dynamics. *arXiv preprint arXiv:1501.00438*, 2015.

[174] Z. Wang, S. Mohamed, and D. Nando. Adaptive Hamiltonian and Riemann manifold Monte Carlo. In *Proceedings of the 30th International Conference on Machine Learning*, pages 1462–1470, 2013.

[175] M. Welling and Y. W. Teh. Bayesian learning via stochastic gradient Langevin dynamics. In *Proceedings of the 28th International Conference on Machine Learning*, pages 681–688, 2011.

[176] M. Whittle and K. P. Travis. Dynamic simulations of colloids by core-modified dissipative particle dynamics. *The Journal of Chemical Physics*, 132(12):124906, 2010.

[177] H. Yoshida. Construction of higher order symplectic integrators. *Physics Letters A*, 150(5-7):262–268, 1990.

Appendix A

Integration Schemes

We list here detailed integration steps for each momentum-conserving method and methods for noisy gradient systems. A fixed stepsize h is used unless otherwise stated. Verlet neighbor lists [172] are used throughout each method .

DPD Velocity-Verlet: DPD-VV

For each particle i,

$$\begin{aligned}
\mathbf{p}_i^{n+1/2} &= \mathbf{p}_i^n + \left(h\mathbf{F}_i^{\mathrm{C}}(\mathbf{q}^n) + h\mathbf{F}_i^{\mathrm{D}}(\mathbf{q}^n, \mathbf{p}^n) + \sqrt{h}\mathbf{F}_i^{\mathrm{R}}(\mathbf{q}^n)\right)/2\,, \\
\mathbf{q}_i^{n+1} &= \mathbf{q}_i^n + hm_i^{-1}\mathbf{p}_i^{n+1/2}\,, \\
\mathbf{p}_i^{n+1} &= \mathbf{p}_i^{n+1/2} + \left(h\mathbf{F}_i^{\mathrm{C}}(\mathbf{q}^{n+1}) + h\mathbf{F}_i^{\mathrm{D}}(\mathbf{q}^{n+1}, \mathbf{p}^{n+1/2}) + \sqrt{h}\mathbf{F}_i^{\mathrm{R}}(\mathbf{q}^{n+1})\right)/2\,,
\end{aligned}$$

where $\mathbf{F}_i^{\mathrm{C}}(\mathbf{q})$, $\mathbf{F}_i^{\mathrm{D}}(\mathbf{q}, \mathbf{p})$ and $\mathbf{F}_i^{\mathrm{R}}(\mathbf{q})$ are conservative, dissipative and random forces, respectively, in the standard DPD system. Note that, at the end of each integration step, the dissipative forces $\mathbf{F}_i^{\mathrm{D}}(\mathbf{q}^{n+1}, \mathbf{p}^{n+1})$ are further updated by using the up-to-date velocities (momenta).

Shardlow's Splitting Method: DPD-S1

For each interacting pair within cutoff radius ($q_{ij} < r_{\mathrm{c}}$),

$$\begin{aligned}
\mathbf{p}_i^{n+1/4} &= \mathbf{p}_i^n - \gamma\omega^{\mathrm{D}}(q_{ij}^n)(\hat{\mathbf{q}}_{ij}^n \cdot \mathbf{v}_{ij}^n)\hat{\mathbf{q}}_{ij}^n h/2 + \sigma\omega^{\mathrm{R}}(q_{ij}^n)\hat{\mathbf{q}}_{ij}^n\sqrt{h}\mathrm{R}_{ij}^n/2\,,\\
\mathbf{p}_j^{n+1/4} &= \mathbf{p}_j^n + \gamma\omega^{\mathrm{D}}(q_{ij}^n)(\hat{\mathbf{q}}_{ij}^n \cdot \mathbf{v}_{ij}^n)\hat{\mathbf{q}}_{ij}^n h/2 - \sigma\omega^{\mathrm{R}}(q_{ij}^n)\hat{\mathbf{q}}_{ij}^n\sqrt{h}\mathrm{R}_{ij}^n/2\,,\\
\mathbf{p}_i^{n+2/4} &= \mathbf{p}_i^{n+1/4} + \sigma\omega^{\mathrm{R}}(q_{ij}^n)\hat{\mathbf{q}}_{ij}^n\sqrt{h}\mathrm{R}_{ij}^n/2\\
&\quad - \frac{\gamma\omega^{\mathrm{D}}(q_{ij}^n)h}{2(1+\gamma\omega^{\mathrm{D}}(q_{ij}^n)h)}\left((\hat{\mathbf{q}}_{ij}^n \cdot \mathbf{v}_{ij}^{n+1/4})\hat{\mathbf{q}}_{ij}^n + \sigma\omega^{\mathrm{R}}(q_{ij}^n)\hat{\mathbf{q}}_{ij}^n\sqrt{h}\mathrm{R}_{ij}^n\right)\,,\\
\mathbf{p}_j^{n+2/4} &= \mathbf{p}_j^{n+1/4} - \sigma\omega^{\mathrm{R}}(q_{ij}^n)\hat{\mathbf{q}}_{ij}^n\sqrt{h}\mathrm{R}_{ij}^n/2\\
&\quad + \frac{\gamma\omega^{\mathrm{D}}(q_{ij}^n)h}{2(1+\gamma\omega^{\mathrm{D}}(q_{ij}^n)h)}\left((\hat{\mathbf{q}}_{ij}^n \cdot \mathbf{v}_{ij}^{n+1/4})\hat{\mathbf{q}}_{ij}^n + \sigma\omega^{\mathrm{R}}(q_{ij}^n)\hat{\mathbf{q}}_{ij}^n\sqrt{h}\mathrm{R}_{ij}^n\right)\,,
\end{aligned}$$

where R_{ij}^n are normally distributed variables with zero mean and unit variance.

For each particle i,

$$\begin{aligned}
\mathbf{p}_i^{n+3/4} &= \mathbf{p}_i^{n+2/4} + h\mathbf{F}_i^{\mathrm{C}}(\mathbf{q}^n)/2\,,\\
\mathbf{q}_i^{n+1} &= \mathbf{q}_i^n + hm_i^{-1}\mathbf{p}_i^{n+3/4}\,,\\
\mathbf{p}_i^{n+1} &= \mathbf{p}_i^{n+3/4} + h\mathbf{F}_i^{\mathrm{C}}(\mathbf{q}^{n+1})/2\,.
\end{aligned}$$

DPD-Trotter Scheme: DPD-Trotter

For each interacting pair within cutoff radius ($q_{ij} < r_{\mathrm{c}}$),

$$\begin{aligned}
\mathbf{p}_i^{n+1/2} &= \mathbf{p}_i^n + m_{ij}\Delta v_{ij}(\mathbf{q}^n, \mathbf{p}^n)\hat{\mathbf{q}}_{ij}^n\,,\\
\mathbf{p}_j^{n+1/2} &= \mathbf{p}_j^n - m_{ij}\Delta v_{ij}(\mathbf{q}^n, \mathbf{p}^n)\hat{\mathbf{q}}_{ij}^n\,,
\end{aligned}$$

with

$$\Delta v_{ij} = \left[\hat{\mathbf{q}}_{ij} \cdot \mathbf{v}_{ij} - \hat{\mathbf{q}}_{ij} \cdot \mathbf{F}_{ij}^{\mathrm{C}}(\mathbf{q})/(\tau m_{ij})\right](e^{-\tau h/2} - 1) + \sqrt{k_{\mathrm{B}}T(1-e^{-\tau h})/m_{ij}}\mathrm{R}_{ij}\,,$$

where $\tau = \gamma\omega^{\mathrm{D}}/m_{ij}$, $m_{ij} = m_i m_j/(m_i + m_j)$ and R_{ij} are normally distributed variables with zero mean and unit variance.

For each particle i,

$$\mathbf{q}_i^{n+1} = \mathbf{q}_i^n + hm_i^{-1}\mathbf{p}_i^{n+1/2}\,.$$

For each interacting pair within cutoff radius ($q_{ij} < r_{\mathrm{c}}$),

$$\begin{aligned}\mathbf{p}_i^{n+1} &= \mathbf{p}_i^{n+1/2} + m_{ij}\Delta v_{ij}(\mathbf{q}^{n+1}, \mathbf{p}^{n+1/2})\hat{\mathbf{q}}_{ij}^{n+1}\,,\\ \mathbf{p}_j^{n+1} &= \mathbf{p}_j^{n+1/2} - m_{ij}\Delta v_{ij}(\mathbf{q}^{n+1}, \mathbf{p}^{n+1/2})\hat{\mathbf{q}}_{ij}^{n+1}\,.\end{aligned}$$

Lowe–Andersen Thermostat: LA

For each particle i,

$$\begin{aligned}\mathbf{p}_i^{n+1/3} &= \mathbf{p}_i^{n} + h\mathbf{F}_i^{\mathrm{C}}(\mathbf{q}^{n})/2\,,\\ \mathbf{q}_i^{n+1} &= \mathbf{q}_i^{n} + hm_i^{-1}\mathbf{p}_i^{n+1/3}\,,\\ \mathbf{p}_i^{n+2/3} &= \mathbf{p}_i^{n+1/3} + h\mathbf{F}_i^{\mathrm{C}}(\mathbf{q}^{n+1})/2\,.\end{aligned}$$

For each interacting pair within cutoff radius ($q_{ij} < r_{\mathrm{c}}$), with probability $P = \Gamma h$,

$$\begin{aligned}\mathbf{p}_i^{n+1} &= \mathbf{p}_i^{n+2/3} + \Delta\mathbf{p}_{ij}\,,\\ \mathbf{p}_j^{n+1} &= \mathbf{p}_j^{n+2/3} - \Delta\mathbf{p}_{ij}\,,\end{aligned}$$

where

$$\Delta\mathbf{p}_{ij} = m_{ij}\left[\mathrm{R}_{ij}\sqrt{k_{\mathrm{B}}T/m_{ij}} - \hat{\mathbf{q}}_{ij}^{n+1}\cdot\mathbf{v}_{ij}^{n+2/3}\right]\hat{\mathbf{q}}_{ij}^{n+1}\,.$$

Peters Scheme II: Peters

For each particle i,

$$\begin{aligned}\mathbf{p}_i^{n+1/3} &= \mathbf{p}_i^{n} + h\mathbf{F}_i^{\mathrm{C}}(\mathbf{q}^{n})/2\,,\\ \mathbf{q}_i^{n+1} &= \mathbf{q}_i^{n} + hm_i^{-1}\mathbf{p}_i^{n+1/3}\,,\\ \mathbf{p}_i^{n+2/3} &= \mathbf{p}_i^{n+1/3} + h\mathbf{F}_i^{\mathrm{C}}(\mathbf{q}^{n+1})/2\,.\end{aligned}$$

For each interacting pair within cutoff radius ($q_{ij} < r_{\mathrm{c}}$),

$$\begin{aligned}\mathbf{p}_i^{n+1} &= \mathbf{p}_i^{n+2/3} + \Delta\mathbf{p}_{ij}\,,\\ \mathbf{p}_j^{n+1} &= \mathbf{p}_j^{n+2/3} - \Delta\mathbf{p}_{ij}\,,\end{aligned}$$

with

$$\Delta\mathbf{p}_{ij} = \left[-\gamma_{ij}(\hat{\mathbf{q}}_{ij}^{n+1}\cdot\mathbf{v}_{ij}^{n+2/3})h + \sigma_{ij}\sqrt{h}\mathrm{R}_{ij}\right]\hat{\mathbf{q}}_{ij}^{n+1}\,,$$

where

$$\gamma_{ij} = \frac{m_{ij}}{h}\left(1 - \exp\left[-\frac{\gamma\omega^{\mathrm{D}}(q_{ij})h}{m_{ij}}\right]\right), \quad \sigma_{ij} = \frac{k_{\mathrm{B}}Tm_{ij}}{h}\left(1 - \exp\left[-\frac{2\gamma\omega^{\mathrm{D}}(q_{ij})h}{m_{ij}}\right]\right),$$

where γ and σ are the same dissipation and random coefficients respectively as in the standard DPD system.

Nosé–Hoover–Lowe–Andersen Thermostat: NHLA

For each particle i,

$$\begin{aligned}\mathbf{q}_i^{n+1} &= \mathbf{q}_i^n + h\mathbf{v}_i^n + h^2\mathbf{F}_i^{\mathrm{C}}(\mathbf{q}^n)/2\,,\\ \mathbf{v}_i^{n+1/4} &= \mathbf{v}_i^n + h\mathbf{F}_i^{\mathrm{C}}(\mathbf{q}^n)/2\,.\end{aligned}$$

For $(1-P)$ fraction interacting pairs within cutoff radius ($q_{ij} < r_{\mathrm{c}}$),

$$\begin{aligned}\mathbf{F}_i^{\mathrm{D}}(\mathbf{q}^{n+1},\mathbf{p}^{n+1/4}) &= \mathbf{F}_i^{\mathrm{D}}(\mathbf{q}^n,\mathbf{p}^n) + \mathbf{F}_{ij}^{\mathrm{D}}(\mathbf{q}^{n+1},\mathbf{p}^{n+1/4})\,,\\ \mathbf{F}_j^{\mathrm{D}}(\mathbf{q}^{n+1},\mathbf{p}^{n+1/4}) &= \mathbf{F}_j^{\mathrm{D}}(\mathbf{q}^n,\mathbf{p}^n) - \mathbf{F}_{ij}^{\mathrm{D}}(\mathbf{q}^{n+1},\mathbf{p}^{n+1/4})\,,\end{aligned}$$

where

$$\mathbf{F}_{ij}^{\mathrm{D}}(\mathbf{q}^{n+1},\mathbf{p}^{n+1/4}) = \alpha\omega^{\mathrm{R}}(q_{ij})(\hat{\mathbf{q}}_{ij}^{n+1}\cdot\mathbf{v}_{ij}^{n+1/4})\hat{\mathbf{q}}_{ij}^{n+1}\,,$$

where α is a coupling parameter chosen as $0.9/(\rho h)$, and ρ is the particle density.

For each particle i,

$$\begin{aligned}\mathbf{v}_i^{n+2/4} &= \mathbf{v}_i^{n+1/4} + h\mathbf{F}_i^{\mathrm{C}}(\mathbf{q}^{n+1})/2\,,\\ \mathbf{p}_i^{n+3/4} &= \mathbf{p}_i^{n+2/4} + h\left(1 - \tilde{T}_{\mathrm{k}}/T_0\right)\mathbf{F}_i^{\mathrm{D}}(\mathbf{q}^{n+1},\mathbf{p}^{n+1/4})\,,\end{aligned}$$

where $\tilde{T}_{\mathrm{k}}$ is the momentary kinetic temperature and T_0 is the desired temperature.

For the remaining (P) fraction interacting pairs within cutoff radius ($q_{ij} < r_{\mathrm{c}}$),

$$\begin{aligned}\mathbf{p}_i^{n+1} &= \mathbf{p}_i^{n+3/4} + \Delta\mathbf{p}_{ij}\,,\\ \mathbf{p}_j^{n+1} &= \mathbf{p}_j^{n+3/4} - \Delta\mathbf{p}_{ij}\,,\end{aligned}$$

where

$$\Delta\mathbf{p}_{ij} = m_{ij}\left[\mathrm{R}_{ij}\sqrt{k_{\mathrm{B}}T/m_{ij}} - \hat{\mathbf{q}}_{ij}^{n+1}\cdot\mathbf{v}_{ij}^{n+3/4}\right]\hat{\mathbf{q}}_{ij}^{n+1}\,.$$

Pairwise Nosé–Hoover Thermostat: PNH

For each particle i,

$$\begin{aligned}
\mathbf{p}_i^{n+1/2} &= \mathbf{p}_i^n + h\left(\mathbf{F}_i^{\mathrm{C}}(\mathbf{q}^n) - \xi^n \mathbf{V}_i(\mathbf{q}^n, \mathbf{p}^{n-1/2})\right)/2\,, \\
\xi^{n+1/2} &= \xi^n + hG(\mathbf{q}^n, \mathbf{p}^{n-1/2})/2, \\
\mathbf{q}_i^{n+1} &= \mathbf{q}_i^n + h m_i^{-1} \mathbf{p}_i^{n+1/2}\,, \\
\mathbf{p}_i^{n+1} &= \mathbf{p}_i^{n+1/2} + h\left(\mathbf{F}_i^{\mathrm{C}}(\mathbf{q}^{n+1}) - \xi^{n+1/2} \mathbf{V}_i(\mathbf{q}^{n+1}, \mathbf{p}^{n+1/2})\right)/2\,, \\
\xi^{n+1} &= \xi^{n+1/2} + hG(\mathbf{q}^{n+1}, \mathbf{p}^{n+1/2})/2\,,
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{V}_i(\mathbf{q}, \mathbf{p}) &= \sum_{j \neq i} \omega^{\mathrm{D}}(q_{ij})(\hat{\mathbf{q}}_{ij} \cdot \mathbf{v}_{ij})\hat{\mathbf{q}}_{ij}\,, \\
G(\mathbf{q}, \mathbf{p}) &= \mu^{-1} \sum_i \sum_{j>i} \omega^{\mathrm{D}}(q_{ij}) \left[(\mathbf{v}_{ij} \cdot \hat{\mathbf{q}}_{ij})^2 - k_{\mathrm{B}}T/m_{ij}\right]\,.
\end{aligned}$$

Symmetric Pairwise Nosé–Hoover–Langevin Thermostat: PNHL-S

For each particle i,

$$\begin{aligned}
\mathbf{q}_i^{n+1/2} &= \mathbf{q}_i^n + h m_i^{-1} \mathbf{p}_i^n/2\,, \\
\mathbf{p}_i^{n+1/4} &= \mathbf{p}_i^n + h\mathbf{F}_i^{\mathrm{C}}(\mathbf{q}^{n+1/2})/2\,.
\end{aligned}$$

For each interacting pair within cutoff radius ($q_{ij} < r_{\mathrm{c}}$),

$$\begin{aligned}
\mathbf{p}_i^{n+2/4} &= \mathbf{p}_i^{n+1/4} + m_{ij}\Delta v_{ij}(\mathbf{q}^{n+1/2}, \mathbf{p}^{n+1/4}, \xi^n)\hat{\mathbf{q}}_{ij}^{n+1/2}\,, \\
\mathbf{p}_j^{n+2/4} &= \mathbf{p}_j^{n+1/4} - m_{ij}\Delta v_{ij}(\mathbf{q}^{n+1/2}, \mathbf{p}^{n+1/4}, \xi^n)\hat{\mathbf{q}}_{ij}^{n+1/2}\,,
\end{aligned}$$

where

$$\Delta v_{ij} = (\hat{\mathbf{q}}_{ij} \cdot \mathbf{v}_{ij})\left(\exp(-\xi\omega^{\mathrm{D}}(h/2)/m_{ij}) - 1\right)\,.$$

For additional variable ξ,

$$\begin{aligned}
\xi^{n+1/3} &= \xi^n + hG(\mathbf{q}^{n+1/2}, \mathbf{p}^{n+2/4})/2\,, \\
\xi^{n+2/3} &= e^{-\tilde{\gamma}h}\xi^{n+1/3} + \sqrt{k_{\mathrm{B}}T(1 - e^{-2\tilde{\gamma}h})/\mu}\mathrm{R}^n\,, \\
\xi^{n+1} &= \xi^{n+2/3} + hG(\mathbf{q}^{n+1/2}, \mathbf{p}^{n+2/4})/2\,,
\end{aligned}$$

where R^n are normally distributed variables with zero mean and unit variance.

For each interacting pair within cutoff radius ($q_{ij} < r_\mathrm{c}$),

$$\begin{aligned}
\mathbf{p}_i^{n+3/4} &= \mathbf{p}_i^{n+2/4} + m_{ij}\Delta v_{ij}(\mathbf{q}^{n+1/2}, \mathbf{p}^{n+2/4}, \xi^{n+1})\hat{\mathbf{q}}_{ij}^{n+1/2}\,, \\
\mathbf{p}_j^{n+3/4} &= \mathbf{p}_j^{n+2/4} - m_{ij}\Delta v_{ij}(\mathbf{q}^{n+1/2}, \mathbf{p}^{n+2/4}, \xi^{n+1})\hat{\mathbf{q}}_{ij}^{n+1/2}\,.
\end{aligned}$$

For each particle i,

$$\begin{aligned}
\mathbf{p}_i^{n+1} &= \mathbf{p}_i^{n+3/4} + h\mathbf{F}_i^\mathrm{C}(\mathbf{q}^{n+1/2})/2\,, \\
\mathbf{q}_i^{n+1} &= \mathbf{q}_i^{n+1/2} + hm_i^{-1}\mathbf{p}_i^{n+1}/2\,.
\end{aligned}$$

Nonsymmetric Pairwise Nosé–Hoover–Langevin Thermostat: PNHL-N

For each particle i,

$$\begin{aligned}
\mathbf{q}_i^{n+1/2} &= \mathbf{q}_i^{n} + hm_i^{-1}\mathbf{p}_i^{n}/2\,, \\
\mathbf{p}_i^{n+1/4} &= \mathbf{p}_i^{n} + h\mathbf{F}_i^\mathrm{C}(\mathbf{q}^{n+1/2})/2\,.
\end{aligned}$$

For each interacting pair within cutoff radius ($q_{ij} < r_\mathrm{c}$),

$$\begin{aligned}
\mathbf{p}_i^{n+2/4} &= \mathbf{p}_i^{n+1/4} + m_{ij}\Delta v_{ij}(\mathbf{q}^{n+1/2}, \mathbf{p}^{n+1/4}, \xi^{n})\hat{\mathbf{q}}_{ij}^{n+1/2}\,, \\
\mathbf{p}_j^{n+2/4} &= \mathbf{p}_j^{n+1/4} - m_{ij}\Delta v_{ij}(\mathbf{q}^{n+1/2}, \mathbf{p}^{n+1/4}, \xi^{n})\hat{\mathbf{q}}_{ij}^{n+1/2}\,,
\end{aligned}$$

where

$$\Delta v_{ij} = (\hat{\mathbf{q}}_{ij} \cdot \mathbf{v}_{ij}) \left(\exp(-\xi\omega^\mathrm{D}(h/2)/m_{ij}) - 1\right)\,.$$

For additional variable ξ,

$$\begin{aligned}
\xi^{n+1/3} &= \xi^{n} + hG(\mathbf{q}^{n+1/2}, \mathbf{p}^{n+2/4})/2\,, \\
\xi^{n+2/3} &= e^{-\tilde{\gamma}h}\xi^{n+1/3} + \sqrt{k_\mathrm{B}T(1 - e^{-2\tilde{\gamma}h})/\mu}\mathrm{R}^n\,, \\
\xi^{n+1} &= \xi^{n+2/3} + hG(\mathbf{q}^{n+1/2}, \mathbf{p}^{n+2/4})/2\,.
\end{aligned}$$

For each interacting pair within cutoff radius ($q_{ij} < r_\mathrm{c}$),

$$\begin{aligned}
\mathbf{p}_i^{n+3/4} &= \mathbf{p}_i^{n+2/4} + m_{ij}\Delta v_{ij}(\mathbf{q}^{n+1/2}, \mathbf{p}^{n+2/4}, \xi^{n+1})\hat{\mathbf{q}}_{ij}^{n+1/2}\,, \\
\mathbf{p}_j^{n+3/4} &= \mathbf{p}_j^{n+2/4} - m_{ij}\Delta v_{ij}(\mathbf{q}^{n+1/2}, \mathbf{p}^{n+2/4}, \xi^{n+1})\hat{\mathbf{q}}_{ij}^{n+1/2}\,.
\end{aligned}$$

For each particle i,

$$\mathbf{q}_i^{n+1} = \mathbf{q}_i^{n+1/2} + hm_i^{-1}\mathbf{p}_i^{n+3/4}/2\,,$$
$$\mathbf{p}_i^{n+1} = \mathbf{p}_i^{n+3/4} + h\mathbf{F}_i^{\mathrm{C}}(\mathbf{q}^{n+1})/2\,.$$

Pairwise Adaptive Langevin Thermostat: PAdL

For each particle i,

$$\mathbf{q}_i^{n+1/2} = \mathbf{q}_i^{n} + hm_i^{-1}\mathbf{p}_i^{n}/2\,,$$
$$\mathbf{p}_i^{n+1/4} = \mathbf{p}_i^{n} + h\mathbf{F}_i^{\mathrm{C}}(\mathbf{q}^{n+1/2})/2\,.$$

For each interacting pair within cutoff radius ($q_{ij} < r_{\mathrm{c}}$),

$$\mathbf{p}_i^{n+2/4} = \mathbf{p}_i^{n+1/4} + m_{ij}\Delta v_{ij}(\mathbf{q}^{n+1/2}, \mathbf{p}^{n+1/4}, \xi^n)\hat{\mathbf{q}}_{ij}^{n+1/2}\,,$$
$$\mathbf{p}_j^{n+2/4} = \mathbf{p}_j^{n+1/4} - m_{ij}\Delta v_{ij}(\mathbf{q}^{n+1/2}, \mathbf{p}^{n+1/4}, \xi^n)\hat{\mathbf{q}}_{ij}^{n+1/2}\,,$$

where,
if ($\xi^n \neq 0$):

$$\Delta v_{ij} = (\hat{\mathbf{q}}_{ij}\cdot\mathbf{v}_{ij})\left(\exp(-\tilde{\tau}h/2) - 1\right) + \sigma\sqrt{\frac{1-\exp(-\tilde{\tau}h)}{2\xi^n m_{ij}}}\mathrm{R}_{ij}\,,$$

where $\tilde{\tau} = \xi\omega^{\mathrm{D}}/m_{ij}$, $m_{ij} = m_i m_j/(m_i + m_j)$ and R_{ij} are normally distributed variables with zero mean and unit variance.

else:

$$\Delta v_{ij} = \sigma(\omega^{\mathrm{R}}/m_{ij})\sqrt{h/2}\mathrm{R}_{ij}\,.$$

For additional variable ξ,

$$\xi^{n+1} = \xi^n + hG(\mathbf{q}^{n+1/2}, \mathbf{p}^{n+2/4})/2\,.$$

For each interacting pair within cutoff radius ($q_{ij} < r_{\mathrm{c}}$),

$$\mathbf{p}_i^{n+3/4} = \mathbf{p}_i^{n+2/4} + m_{ij}\Delta v_{ij}(\mathbf{q}^{n+1/2}, \mathbf{p}^{n+2/4}, \xi^{n+1})\hat{\mathbf{q}}_{ij}^{n+1/2}\,,$$
$$\mathbf{p}_j^{n+3/4} = \mathbf{p}_j^{n+2/4} - m_{ij}\Delta v_{ij}(\mathbf{q}^{n+1/2}, \mathbf{p}^{n+2/4}, \xi^{n+1})\hat{\mathbf{q}}_{ij}^{n+1/2}\,,$$

where,
if ($\xi^{n+1} \neq 0$):

$$\Delta v_{ij} = (\hat{\mathbf{q}}_{ij}\cdot\mathbf{v}_{ij})\left(\exp(-\tilde{\tau}h/2) - 1\right) + \sigma\sqrt{\frac{1-\exp(-\tilde{\tau}h)}{2\xi^{n+1} m_{ij}}}\mathrm{R}_{ij}\,.$$

else:

$$\Delta v_{ij} = \sigma(\omega^{\mathrm{R}}/m_{ij})\sqrt{h/2}\mathrm{R}_{ij}\,.$$

For each particle i,

$$\begin{aligned}\mathbf{p}_i^{n+1} &= \mathbf{p}_i^{n+3/4} + h\mathbf{F}_i^{\mathrm{C}}(\mathbf{q}^{n+1/2})/2\,,\\ \mathbf{q}_i^{n+1} &= \mathbf{q}_i^{n+1/2} + hm_i^{-1}\mathbf{p}_i^{n+1}/2\,.\end{aligned}$$

Stochastic Gradient Langevin Dynamics: SGLD

$$\mathbf{q}_{n+1} = \mathbf{q}_n + h_n\tilde{\mathbf{F}}(\mathbf{q}_n) + \sqrt{2\beta^{-1}h_n}\mathbf{R}_n\,,$$

where h_n is a sequence of stepsizes decreasing to zero, $\tilde{\mathbf{F}}(\mathbf{q}_n)$ is the noisy force (5.15) and $\mathbf{R}_n$ is a vector of i.i.d. standard normal random variables.

Modified Stochastic Gradient Langevin Dynamics: mSGLD

$$\mathbf{q}_{n+1} = \mathbf{q}_n + h\tilde{\mathbf{F}}(\mathbf{q}_n) + \sqrt{2\beta^{-1}h}\left(\mathbf{I} - \frac{h}{4}\mathrm{Cov}\tilde{\mathbf{F}}(\mathbf{q}_n)\right)\mathbf{R}_n\,,$$

where $\mathrm{Cov}\tilde{\mathbf{F}}(\mathbf{q}_n)$ is the covariance matrix of the noisy force.

Nonsymmetric Stochastic Gradient Nosé–Hoover Thermostat: SGNHT-N

$$\begin{aligned}\mathbf{p}_{n+1} &= \mathbf{p}_n + h\tilde{\mathbf{F}}(\mathbf{q}_n) - h\xi_n\mathbf{p}_n + \sqrt{h}\sigma_{\mathrm{A}}\mathbf{M}^{1/2}\mathbf{R}_n\,,\\ \mathbf{q}_{n+1} &= \mathbf{q}_n + h\mathbf{M}^{-1}\mathbf{p}_{n+1}\,,\\ \xi_{n+1} &= \xi_n + h\mu^{-1}\left(\mathbf{p}_{n+1}^T\mathbf{M}^{-1}\mathbf{p}_{n+1} - N_{\mathrm{d}}k_{\mathrm{B}}T\right)\,.\end{aligned}$$

Symmetric Stochastic Gradient Nosé–Hoover Thermostat: SGNHT-S

$$
\begin{aligned}
\mathbf{p}_{n+1/3} &= \mathbf{p}_n + (h/2)\tilde{\mathbf{F}}(\mathbf{q}_n)\,, \\
\mathbf{q}_{n+1/2} &= \mathbf{q}_n + (h/2)\mathbf{M}^{-1}\mathbf{p}_{n+1/3}\,, \\
\xi_{n+1/2} &= \xi_n + (h/2)\mu^{-1}\left(\mathbf{p}_{n+1/3}^T\mathbf{M}^{-1}\mathbf{p}_{n+1/3} - N_{\mathrm{d}}k_{\mathrm{B}}T\right)\,, \\
\text{if } (\xi_{n+1/2} \neq 0): \quad \mathbf{p}_{n+2/3} &= e^{-\xi_{n+1/2}h}\mathbf{p}_{n+1/3} + \sigma_{\mathrm{A}}\sqrt{(1-e^{-2\xi_{n+1/2}h})/(2\xi_{n+1/2})}\mathbf{M}^{1/2}\mathbf{R}_n\,, \\
\text{else}: \quad \mathbf{p}_{n+2/3} &= \mathbf{p}_{n+1/3} + \sqrt{h}\sigma_{\mathrm{A}}\mathbf{M}^{1/2}\mathbf{R}_n\,, \\
\xi_{n+1} &= \xi_{n+1/2} + (h/2)\mu^{-1}\left(\mathbf{p}_{n+2/3}^T\mathbf{M}^{-1}\mathbf{p}_{n+2/3} - N_{\mathrm{d}}k_{\mathrm{B}}T\right)\,, \\
\mathbf{q}_{n+1} &= \mathbf{q}_{n+1/2} + (h/2)\mathbf{M}^{-1}\mathbf{p}_{n+2/3}\,, \\
\mathbf{p}_{n+1} &= \mathbf{p}_{n+2/3} + (h/2)\tilde{\mathbf{F}}(\mathbf{q}_{n+1})\,.
\end{aligned}
$$